# Oppgaveutvikling til PISA

Hovedoppgave i realfagdidaktikk

av

Aleksandr Rødsten


Institutt for lærerutdanning og skoleutvikling

Universitetet i Oslo

Mai 2005

# Forord

Etter mange år i skolen, og med de forandringer Reform 94 medførte, følte jeg behov for faglig påfyll. Jeg gjennomførte og tok eksamen i Allmenn Statistikk, ST 001. Ønsket om å lære mer ble ikke svekket.

Jeg vil i den forbindelse takke Svein Sjøberg, som både gjennom sin undervisning og samtaler styrket min lyst til videre studier.

I samråd med Gunnar Gjone kom jeg fram til at en hovedoppgave knyttet til utvikling og analyse av oppgaver måtte være interessant. Siden PISA 2003 var i oppstartfasen var det naturlig å knytte arbeidet opp mot dette. Siden jeg hele tiden har vært i fullt arbeid som lærer, har det tatt noe tid å gjennomføre studiet. Jeg kan derfor først nå, etter at PISA 2003 er sluttført, legge fram mine resultater, og håper de kan være av interesse for flere.

Jeg vil takke mine to veileder, Gunnar Gjone og Svein Lie for gode råd, samtaler og inspirasjon. Jeg vil også takke Marion Caspersen for hennes hjelp og støtte.

Oslo, mai 2005

Aleksandr Rødsten

# Innhold

# 1. Innledning

Jeg arbeider som lærer i Fengselsundervisningen i Oslo (Grønland Voksenopplæringssenter). Herunder hører undervisningen ved Bredtveit fengsel, forvarings- og sikringsanstalt for kvinner, hvor jeg har faglig ansvar for matematikk på videregående nivå. Her tilbys tilrettelagt undervisning, og det tas opp nye elever når som helst i skoleåret. Siden elever i samme undervisningsgruppe er på forskjellige nivåer, er det naturlig og nødvendig å differensiere innenfor gruppens rammer. Mange har en avkortet ungdomsskole bak seg og store huller i kunnskapene. Derfor vil bruken av tradisjonelle lærebøker og oppgavesamlinger tilpasset for hele gruppen, være vanskelig. I stedet må undervisningen tilpasses hver enkelt elev. Jeg har gjennom mange år utviklet oppgaver tilpasset denne undervisningsform. For noen år siden har jeg flere ganger laget eksamensoppgaver etter bestilling fra Rådet for videregående opplæring. Mange av disse oppgavene har senere dukket opp i forskjellige oppgavesamlinger.

Etter mer enn 30 års arbeid i skolen, så jeg behov for å få nye impulser for blant annet å kunne utvikle bedre oppgaver for diagnostisering. Det var da naturlig å tenke på å ta hovedfag i realfagdidaktikk. I den forbindelse ble jeg invitert av Institutt for Lærerutdanning og Skoleutvikling (ILS) til å utvikle oppgaveforslag til PISA-prosjektet 2003. De oppgaveforslagene jeg har utviklet for PISA, har jeg også hatt anledning til å teste ut på jentene på Bredtveit, og på den måten fått noen korrigeringer. Et eksempel på det er fra oppgaven Terninger (spørsmål 32) der en av jentene gjorde meg oppmerksom på enda en ny måte en terning kan konstrueres på, som verken jeg eller andre som hadde gjennomgått eller prøvd seg på oppgaven, hadde funnet.

Ved siden av å utvikle oppgaver til PISA-undersøkelsen, var den psykometriske[1] testingen av besvarelsene en viktig del av min hovedoppgave. Derfor får analysen av besvarelsene en sentral plass.

De 8 oppgavene jeg utviklet til pilottesten, består hver av 4 til 6 spørsmål. PISA satte som en forutsetning at de innsendte oppgaveforslagene på forhånd skulle ha vært testet på en mindre gruppe, i mitt tilfelle ble det 143 elever. Det var ønskelig at oppgavene skulle ta hensyn til den store spredningen i elevenes kunnskaper og evner. Derfor representerer de stor variasjon i vanskegrad.

To av mine oppgaver ble i omarbeidet og forkortet form med i den endelige PISA-undersøkelsen. Det sendes inn et betydelig antall oppgaveforslag fra de deltagende landene. At mine øvrige oppgaver ikke kom med, betyr derfor ikke nødvendigvis at de ikke svarte til forutsetningen. En grundigere analyse av besvarelsene avdekker svakheter ved de enkelte spørsmålene, om de enten skulle omarbeides eller burde fjernes. Analysen vil også gi informasjon om hvorvidt noen av oppgavene vil være bedre egnet i en annen sammenheng.

Ut over å følge retningslinjene fra PISA, hadde jeg også en annen intensjon med oppgaveforslagene. De skulle også ligge innenfor læreplanverket for den 10-årige grunnskolen (L97) i så henseende at de skulle kunne være oppgaver til avgangseksamen for

[1] Psykometri: Anvendelse av måling og statistiske metoder i forbindelse med test.

grunnskolen. Jeg har derfor betraktet felles trekk mellom oppgaver gitt til avgangsprøven og de testoppgaver jeg utviklet for pilottesten. Disse inkluderer også det oppgaveformat (se kapittel 3) som blir benyttet på så forskjellige prøver som pilottesten og en avgangseksamen.

Noen uker etter at testen ble gjennomført, ble avgangsprøven i grunnskolen holdt (2001). Det var interessant å se at noen av oppgavene der hadde klare likhetstrekk med mine oppgaver.

## 1.1 Oppgavens innhold

I kapittel 1 vil jeg presentere problemstillingen for min oppgave, og i den sammenheng gi en kort introduksjon av PISA-prosjektet og dets rammeverk for matematikkfaget.

Ved utviklingen av oppgaveforslag har jeg så langt som mulig, tilstrebet at de ligger innenfor Læreplanverket for den 10-årige grunnskolen (L97). Derfor gir jeg i Kapittel 2 et kort resymé av mål og målområder for matematikkfaget i grunnskolen. Det er naturlig å vurdere mine oppgaveforslag i forhold til de oppgaver som er gitt til avgangseksamen i grunnskolen.

I kapittel 3 berører jeg oppgaveformat. Selv om mine oppgaver i utgangspunktet var utarbeidet for PISA, viste pilottesten at ikke alle var velegnet for den test PISA la opp til. Men oppgavene kunne ha andre anvendelsesområder som for eksempel i en diagnostisk test. Derfor berører jeg også oppgaveformat ved diagnostisering.

Kapittel 4 omhandler metodeteori. Her presenteres de sentrale statistiske begreper og de metoder som ligger til grunn for målingene av elevenes besvarelser. Neste kapittel beskriver oppgaveutviklingen med utgangspunkt i retningslinjer gitt fra PISA. Her presenteres også det kodesystemet jeg benyttet for senere å analysere besvarelsene. I kapittel 6 gis en kort beskrivelse av utprøvingen. Det redegjøres for utvalg av skoler, elevsammensetningen og praktiske sider ved gjennomføringen av pilottesten.

Kapittel 7 er den mest omfattende del av oppgaven. Her blir de seks oppgavene som var med i PISAs egen pilottest, presentert og analysert. I tillegg har jeg også tatt med én av de to oppgavene som ikke ble med i PISAs pilottest: oppgaven *Kast med to terninger*. Ved å følge svarmønsteret mellom flere av spørsmålene i denne oppgaven, kom det fram at denne var lite velegnet til PISA. Hvert av spørsmålene i oppgavene blir statistisk behandlet etter de metoder som er omtalt under kapittel 4. I tillegg gjøres det en oppsummering etter hvert oppgavesett med diskusjon på grunnlag av de funn analysen viser.

Kapittel 8 tar for seg reliabilitetsbetraktninger. Her blir også de spørsmålene som korrelerer dårlig med de øvrige spørsmålene i testen, kommentert.

I kapittel 9 gis en konklusjon på analysen. Her presenteres de funn som er gjort under analysen som svar på problemstillingene presentert i p.1.1.1. Det er også en henvisning til oppsummeringen etter hver oppgave. Til slutt ser jeg på korrelasjonen mellom skårpoengene og den innsats i følge "*innsatstermometeret*" elevene har lagt ned i testen.

I et vedlegg blir den siste oppgaven, *Totallsystemet* presentert med de samme statistiske utregner som i de andre oppgavene. Av flere grunner var denne oppgaven lite velegnet til PISA, og ble da heller ikke med i PISAs pilottest.

### *1.1.1 Oppgavens problemstillinger*

Utgangspunktet for min hovedoppgave var altså å utvikle oppgaver til PISA 2003 der matematikk skulle være i fokus. Disse oppgavene skulle utvikles etter retningslinjer fra ***Mathematics item development for PISA 2003*** (se p. 5.1). Jeg ville også samtidig påse at oppgavene lå innenfor Læreplanverket (L97). Det er da naturlig å sette de to hovedproblemstillingene rundt følgende:

1. **Fungerer mine oppgaveforslag bra psykometrisk og oppfyller de kravene i rammeverket til PISA?**
2. **Ligger mine oppgaveforslag innenfor Læreverket til den 10-årige grunnskolen?**

Ved oppgaveutvikling vil en analyse av besvarelsene gi viktig informasjon uavhengig av hvilken funksjon oppgaven skal ha. Det bør være slik at de som svarte riktig på én oppgave, også gjorde det bedre på hele testen enn de som ikke svarte riktig. En analyse av testbesvarelsene vil da fastslå hvilke oppgaver som svarte til denne forutsetningen. Til en slik test er ensifrede koder (se p.5.3) på besvarelsene tilstrekkelig.

Det kan ofte være vel så interessant å registrere de forskjellige typer feil som fremkommer som de riktige svarene. Dersom analysen viser at en oppgave ikke svarer til de ønsker PISA satte, vil svarmønstret mellom flere spørsmål i et oppgaveforslag klargjøre eventuelle svakheter. Jeg valgte derfor å benytte tosifrede koder for både å registrere de forskjellige typer feil og så å følge svarmønstret mellom spørsmålene. På den måten ville analyse av besvarelsene avdekke andre sider ved oppgavene. En problemstilling vil da være:

3. **Vil noen av oppgavene fungere som en diagnostisk test?**

Det kan være flere årsaker til at en elev ikke besvarer en oppgave i en test. Oppgavens kompleksitet kan føre til at den fremstår som ”vanskelig” for eleven. Tidsnød eller manglende motivasjon, kan også være en årsak. Hvilken plassering oppgaven har i en test, kan også ha innvirkning på svarprosenten. I den neste problemstillingen vil jeg se på om oppgavens tema også kan ha en betydning.

I en oppgave med autentisk kontekst hvor problemstillingen presenteres i tekstform skal eleven benytte sine matematikkunnskaper for å løse problemet. Elevens motivasjon for å gå i kast med en oppgave kan ofte være avhengig av hvilken kontekst problemet presenteres i. Konteksten vil da bli oppfattet som en situasjon lokalisert i en viss distanse fra eleven. Selv om konteksten til en oppgave kan plasseres i en kontinuerlig skala fra eleven avhengig av elevens erfaringsbakgrunn, opererer man i PISA-sammenheng med 5 situasjonsklasser (se p.1.3.3). Det kan være problematisk å fastslå med sikkerhet i hvilken grad oppgavekontekstens avstand fra eleven påvirker svarprosenten. Svarprosenten vil i beste fall fortelle oss om det er en tendens til at den øker ved oppgaver med kontekst ”nær” eleven. Derfor vil det være av interesse å se på følgende problemstilling:

4. **Kan man registrere en tendens til at oppgaver med kontekst ”nær” eleven, øker svarprosenten?**

PISA 2000 viste at det er klare forskjeller mellom jenter og gutters leseferdigheter generelt. (Lie, Kjærnsli, Roe & Turmo 2001, s.95). Det ble i retningslinjer for utvikling av oppgaver, bedt om at ordbruken i oppgavene skal være så enkel som mulig for å ta hensyn til, og begrense effekten av, disse kjønnsforskjeller. Min siste problemstilling blir å registrere slike forskjeller.

5. **Kan eventuelle forskjeller i gutters og jenters prestasjoner forklares med oppgavens formulering?**

# 1.2 PISA-prosjektet

OECD-prosjektet PISA (Programme for International Student Assessment) har som formål å sammenligne 15-åringers kompetanse i lesing, matematikk og naturfag. I motsetning til tidligere internasjonale undersøkelser der hovedsaklig kunnskaper i faget måles, undersøker dette prosjektet også elevens kompetanse på tvers av fagene, som anses å være nødvendig for å fungere som voksne i samfunnet. Elevene skal kunne forstå prosesser, tolke og reflektere over informasjoner og dessuten løse problemer de møter i dagliglivet.

Første fase av PISA-prosjektet startet i 1998 og kuliminerte med undersøkelsen våren 2000 der hovedområdet var lesing. I neste fase, våren 2003, var matematikk hovedområdet. Ca. 2/3 av undersøkelsen er konsentrert om dette faget.

Et viktig formål ved undersøkelsene er å finne årsaker til, og konsekvenser av, observerte prestasjonsforskjeller mellom de deltagende land, for derigjennom å belyse de sterke og svake sidene ved de enkelte lands undervisningssystemer. Derved kan man få kunnskaper som er nødvendige for å utvikle en bedre skole.

## *1.2.1 Mathematical Literacy*

PISA-prosjektets siktemål går langt utover bare å måle 15-åringers kunnskaper og ferdigheter i matematikkfaget i de land som deltar i undersøkelsen. Elevenes praktiske anvendelse av sin matematiske kompetanse er et viktig fokusområde. Dette omfatter ikke bare deres ferdigheter i å mestre matematiske problemstillinger slik de tradisjonelt har møtt i oppgaver presentert i en matematikktime. Tidligere hensikter med presentasjon av matematikkoppgaver, var enten å trene elevene på oppgaveløsninger, eller å vurdere deres kompetanse på avgrensede områder, som for eksempel geometri, algebra eller funksjonslære.

Hensikten med PISA-testen er også å fastslå i hvilken grad elevenes matematikkunnskaper er anvendelig utenfor en undervisningssituasjon. Rammeverket har et uttrykk som skal dekke denne siden av matematikkopplæringen, m*athematical literacy*, som vi på norsk dessverre ikke har en presis oversettelse av. I *Draft framework for the PISA 2003 mathematics assessment*[2] *(January 2001 Draft)* er *mathematical* literacy definert slik:

> *Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to engage in mathematics, in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen. (OECD/PISA 2001a, s.5)*

Parallelt med uttrykket *mathematical literacy* brukes også på engelsk *numeracy* for de grunnleggende matematikkferdigheter som voksne har bruk for i samfunnet. Med dette

[2] Ved utvikling av pilotoppgavene våren 2001 forholdt jeg meg til denne som på det tidspunktet var eneste tilgjengelige versjon.

begrepet menes at man ikke automatisk lærer å bruke matematikk i situasjoner utenfor skolen ved å lære matematikk i skolen. Det vil si at kunnskaper i matematikk ikke automatisk fører til praktisk matematikkunnskap (Wedege 2000, s.6).

I tillegg kommer også fagets betydning for at dagens skoleelever i framtiden skal bli reflekterende samfunnsaktører klart fram. Jeg oppfatter dette også som en forsterkning av skolens plikt til å legge vekt på det allmenndannende aspekt i undervisningen. Det står ikke i motsetning til nytteaspektet ved faget (se kapittel 2).

> *Vi kan altså si at visjonen om at skolen og skolens fag skal fremme dannelse, eller være allmenndannende, går ut på at skolen skal bidra til at elevene utvikler seg til individer som er i stand til å delta på en selvstendig, reflektert og kritisk måte i vårt demokratiske samfunn. (Sjøberg 1998, s.37)*

Vektleggingen av matematikkunnskapenes anvendelse som et redskap som kan brukes i forskjellige situasjoner, blir også sterkt presisert i Læreplanverket for den 10-årige grunnskolen (L97). På dette området er de reformer norsk skole har gjennomført de siste årene, i overensstemmelse med det siktemål PISA skal belyse. I L97 står det:

> *Kunnskaper og ferdigheter i matematikk er et viktig grunnlag for aktiv deltakelse i arbeid og fritid og for å kunne forstå og øve innflytelse på prosesser i samfunnet. Matematikk kan være et hjelpemiddel for å mestre utfordringer for den enkelte. (L97, s.159)*

Et viktig aspekt ved *mathematical literacy* er evnen eleven bør ha tilegnet seg for å kunne sette opp, formulere, løse og tolke løsningen på et matematisk problem. Det gjelder også å kunne anvende forskjellige områder av matematikken for å løse en oppgave enten den matematiske strukturen er klart tilstede eller ikke. Det blir da problemløseren (eleven) som må transformere fram den matematiske strukturen på problemstillingen. Problemstillinger kan spenne fra oppgavetyper som krever bruk av enkle matematiske ferdigheter og begreper, til oppgaver som krever metoder på høyere matematisk nivå.

## 1.3 Rammeverket (Framework)

Rammeverket for PISA framhever tre viktige hovedelementer for de oppgaver som presenteres for undersøkelsen i 2003.

- *sentrale ideer*
- *kompetanseklasser*
- *situasjoner - kontekster*

Disse hovedelementene blir nærmere forklart i de neste kapitlene.

Mer om fellestrekk mellom PISA-prosjektets intensjoner og matematikksynet som fremkommer i L97 blir gjennomgått i p.2.2

### *1.3.1 Sentrale ideer*

Et mål for matematikktesten i PISA-prosjektet, er å evaluere elevenes evne til å anvende matematikk på en måte som øker deres mulighet til aktiv deltakelse i samfunnet. Det er først og fremst det allmennyttige i faget som er gjenstand for testene elevene blir prøvd i. Oppgaveforslagene som er presentert bak, preges derfor av dette.

For å løse en oppgave hentet fra dagliglivet, må elevene bruke kunnskap fra forskjellige områder innenfor matematikken. Det er lite hensiktsmessig å avgrense hver enkel oppgave til hvert av de tradisjonelle emnene som for eksempel algebra, funksjoner eller geometri. Det er heller sjelden at oppgaver med problemer fra dagliglivet kan forstås og løses ved hjelp av bare ett emne fra skolematematikken. Oppgaveforslagene er derfor organisert på en slik måte at løsing av oppgavene krever bruk av kunnskap fra forskjellige matematiske emner. For PISA 2003 har denne organiseringen fått benevnelsen *Sentrale ideer* og er ment å reflektere den utviklingen matematikkfaget har hatt i mange land i de senere årene. Rammeverket er konsentrert omkring følgende fire sentrale ideer: *Kvalitative resonnement, Rom og form, Forandring og sammenheng* og *Usikkerhet*.

**Kvantitativt resonnement**

Ved hjelp av Kvantitativt resonnement forsøker PISA å evaluere om elevene forstår betydningen av tallbegrepet både i forbindelse med relative størrelser, behandling av tall og i elegante matematiske beregninger. Tall og størrelser kan presenteres på forskjellige måter, som gjennom tekst eller diagram. Elevene må vise om de har utviklet ferdigheter til å hente ut og tolke relevant tallinformasjon, slik for eksempel oppgaven *Barnefødsler* (se p.7.3) krever.

En annen viktig side ved *Kvantitativt resonnement*, er at elevene skal vurdere om resultat som fremkommer ved en utregning, er relevant, med andre ord en prøve på om elevene har utviklet rimelige ferdigheter i å estimere, enten ved rask hoderegning eller ved intuitivt å kunne se om resultatet er rimelig i forhold til oppgaven.

**Rom og form**

I dagliglivet opplever elevene estetiske sider ved både natur, kunst og arkitektur og i geometriske figurer de derved møter. Gjennom opplæring får de trening i å oppfatte kjennetegn og egenskaper ved former, mønstre og figurer i rommet. I oppgaveløsingen skal de utføre konstruksjoner og beskrive avbildninger som gir grunnlag for å evaluere forståelsen av hvordan tredimensjonale figurer kan presenteres i to dimensjoner. I oppgaven *Terninger* (se p.7.7) skal elevene markere de forslagene som de tror kan brettes til en terning. Andre egenskaper ved geometriske avbildninger av tredimensjonale konstruksjoner, er parallellforskyvning, speiling og rotasjon. Heri ligger også hvordan perspektiv fungerer når avbildninger av figurer i rommet presenteres i to dimensjoner.

**Forandringer og sammenheng**

Når elevene blir presentert for oppgaver i *Forandring og sammenheng*, hentes tema ofte fra fenomen i omgivelsene. I våre omgivelser finner vi både naturlige og menneskeskapte fenomener. Et eksempel er den naturlige, periodiske endringen som årstidenes temperaturforandringer representerer. Sammenhengen mellom grader målt i Fahrenheit og Celsius, og omregningsformelen mellom disse, er eksempel på et menneskeskapt fenomen. Å løse slike oppgaver krever ofte anvendelse av formler og funksjonsuttrykk som kan være gjengitt i oppgaveteksten. Sammenhengen kan beskrives verbalt, eller formelen som skal anvendes forutsettes kjent av eleven. Ved for eksempel konstruksjon av en trapp vil trinndybden være avhengig av trappehøyden etter trappeformelen (se oppgave *Trappekonstruksjon* p.7.2), som oppgis i teksten i oppgaven.

Det kan også være hensiktsmessig å presentere oppgaver ved tabeller, grafer og/eller geometriske figurer, der eleven leser og selv tolker informasjonen for å løse et problem. Eksempel på dette er oppgaven om *Fibonaccitallene* (se p.7.5) der problemet er å tolke en figur for derigjennom å finne en tallfølge. Hensikten med å bruke forskjellige presentasjonsformer er å prøve elevens evner til å tolke den informasjonen som oppgaven formidler.

**Usikkerhet**

Etter de reformer norsk skole har gjennomgått de siste årene, er *Behandling av data* nå et av målområdene nedfelt i matematikkfaget på ungdomsskolen (se p 2.1). Dette er en utvikling vi deler med mange andre land og PISA-prosjektet dekket derfor dette området i undersøkelsen i 2003.

Dagens 15-åringer blir på mange forskjellige måter konfronter med en mengde informasjoner fra dagliglivet, bl.a. gjennom datainformasjoner, som for eksempel om befolkningstilvekst (se oppgaven *Barnefødsler* p.7.3). De fleste elever har rundt spill med terninger stiftet bekjentskap med sannsynligheter. I oppgaven *Kast med to terninger* (se p.7.6) må elevene regne ut sannsynlighetsutfall.

En viktig side ved *Usikkerhet* er at eleven selv må tolke datainformasjonen og finne ut hva som er relevant for å kunne løse oppgaven, og så gjøre de nødvendige analyser før slutningen trekkes. I dette ligger at eleven ofte må resonnere ut fra empiriske data med en viss usikkerhet der de tall som er nødvendige for å løse oppgaven, må hentes fra en større sammenheng.

### 1.3.2 Kompetanseklasser

> *I PISA tar man ikke utgangspunkt i landenes læreplaner og skolefagenes "pensum", men man tar i hovedsak sikte på å måle elevens evne til aktivt å bruke kunnskaper og erfaringer og hvordan de forholder seg til emner som trolig vil være relevante for framtiden. (Lie, Kjærnsli, Roe & Turmo 2001, s.11)*

I framtiden vil vi ha et betydelig behov for individer som kan håndtere den stigende kompleksitet som vil kjennetegne flere sider av samfunnet. For å håndtere den stigende kompleksiteten, vil det derfor være behov for økt kompetanse på en rekke områder.

Det danske undervisningsministeriet henvendte seg til Mogens Niss som var involvert i PISA-prosjektet med bl.a. spørsmålet:

> *Hvilke matematiske kompetencer skal det være oppbygget hos eleverne på de forskjellige stadier af uddannelsessystemet? (Niss & Jensen 2002)*

Svaret ble at matematisk kompetanse kan oppfattes ved 8 delkompetanser:

1. *Mathematical thinking and reasoning* – som inkluderer både å forstå og behandle omfanget av og grensene for, gitte matematiske konsept.

2. *Mathematical argumentation* – som også omfatter kunnskapen om hva matematiske bevis er og hvordan de skiller seg fra andre matematiske resonnement.

3. *Mathematical communication* – ved også å uttrykke seg på forskjellige måter innenfor områder med matematisk innhold.
4. *Modeling* – å kunne analysere situasjoner og bygge matematiske modeller fra andre situasjoner.
5. *Problem posing and solving* – å kunne formulere og løse matematiske problemer også på flere områder.
6. *Representation* – å kunne beherske forskjellige presentasjonsformer.
7. *Using symbolic, formal and technical language and operations* – å kunne tolke matematiske symboler.
8. *Use of aids and tools* – å være i stand til å nyttegjøre seg forskjellige hjelpemidler for matematisk virksomhet.

(OECD/PISA 2001b s.18,19. Den norske teksten er mine tilføyelser)

De 8 delkompetanser kan naturlig inndeles i to hovedgrupper der 1, 2, 4 og 5 representerer matematisk tenkning rundt spørsmål og svar i, med og om matematikk. Delkompetanse 3, 6, 7 og 8 representerer matematiske presentasjonsformer, språk og redskap.

Kravene til matematikkunnskaper vil for elevene variere både innen utdanning, deres fritid og senere, i arbeidslivet. Det vil da også ligge ulike meninger i hva det vil si å kunne matematikk. Brekke (1995, s.6 -10) peker på *fem komponenter* som utgjør det en kan kalle matematisk kompetanse: 1. *Faktakunnskaper*, 2. *Ferdigheter*, 3. *Begrepsstrukturer*, 4. *Generelle strategier*, 5. *Holdninger*. Disse komponentene danner da også et godt utgangspunkt for tenkningen rundt hva som må være med når det er tale om å utvikle grunnleggende kunnskaper hos elevene.

Vi ser at Brekke (1995) også tar med betydningen av *Holdninger* som en av komponentene for matematisk kompetanse. I en undervisningssituasjon vil dette gjelde både lærerens og elevens syn på hva matematikk er, hvordan undervisningsopplegget er tilrettelagt for den enkelte elev og hvordan eleven møter faget. Økt forventning om å mestre faget, kan ha stor betydning for elevens prestasjon. Tilsvarende vil utfordringer ut over elevens egen kompetanse ofte resultere i redusert innsats eller at eleven gir opp. (Skaalvik 1996, s.83).

Matematikkoppgavene utviklet for PISA klassifiseres i tre kompetanseklasser. Denne klassifiseringen består i hovedsak av at for å kunne løse oppgavene, må elevene benytte forskjellige prosedyrer idet de forskjellige typer kompetanse svarer til forskjellige mentale prosesser. For å løse en matematikkoppgave, må elevene mobilisere sine ferdigheter og den kompetanse de har tilegnet seg gjennom skole og liv. Noen oppgaver krever bruk av enkle matematiske ferdigheter og begreper. For å løse andre oppgavetyper må elevene ”modellere” problemsituasjonen ved å tilpasse, opprette relasjoner mellom, og identifisere forbindelser som en del av den mentale prosessen. Hvordan elevene må gå til verks for å løse en oppgave, er grunnlaget for oppgavenes plassering i kompetanseklasser.

Disse kompetanseklassene er utviklet etter tidligere modeller som er brukt ved mange tidligere undersøkelser i flere land.

Det er ingen klare grenser mellom disse tre kompetanseklassene, men det er opplagt at det kreves større matematisk kompetanse for å løse oppgaver fra kompetanseklasse 3 enn fra de andre kompetanseklassene. For de fleste elever vil oppgaver fra kompetanseklasse 1 virke

”lettere” og derav vil prosenten riktige svar forventes å være høyere. Men det vil alltid være elever med manglende innøvde prosedyrer som også vil oppfatte disse oppgavene som ”vanskelige”, selv om oppgavene stort sett omfatter standardalgoritmer, rutineberegninger, framgangsmåter og problemløsninger som skulle være godt innøvd. Kompetanseklasse 2 vil i tillegg til rutineberegninger også inkludere resonnering, tolkning og refleksjoner. Elevene vil her også måtte vise om de er i stand til å modellere problemet ved å omsette tekstoppgaver til et korrekt matematisk språk for derigjennom å løse problemet.

I kompetanseklasse 3 får elevene presentert oppgaver som krever metoder og kombinasjoner på det høyeste matematiske nivå. Elevene får her ikke de vinkene som oppgaveteksten i de andre kompetanseklassene legger opp til. De må stake ut veien på egen hånd.

Kompetanseklassene kan derved kort beskrives ved sine kjennetegn.

**Kompetanseklasse 1:**

For å løse oppgaver i denne kompetanseklassen må elevene anvende de mest brukte rutinemessige utregninger, elevene blir prøvd i faktakunnskaper og standardalgoritmer. Det forutsettes at elevene husker enkle matematiske framgangsmåter og de får her vist sine tekniske ferdigheter ved å behandle uttrykk som inneholder symboler og formler.

De enkleste matematiske problempresentasjoner hører også med i denne klassen enten det gjelder oppgaver elevene gjenkjenner fra lærebøkene eller oppgaver der teksten gir tilstrekkelig vink om framgangsmåte.

**Kompetanseklasse 2:**

Denne kompetanseklassen inkluderer problemløsninger der elevene må kunne se sammenhengen mellom forskjellige tråder i matematikken. De må kunne kombinere og se helheten i informasjoner for derigjennom å løse oppgaver. I denne kompetanseklassen må elevene velge og utvikle strategier, velge riktige matematiske verktøy og anvende sammensatte metoder i matematiseringen og modellprosessen. Disse ferdighetene reflekterer også elevenes evne til å tolke betydningen av en oppgaveløsning og validiteten av det utførte arbeidet.

**Kompetanseklasse 3:**

I denne kompetanseklassen skal elevene ikke bare matematisere mer sammensatte problem, men også selv utvikle originale løsningsmetoder. Tema for måling i denne klassens kompetanse, bør reflektere elevenes evne til analyse, tolkning, refleksjon omkring og presentasjon av matematiske generaliseringer, argumentasjon og bevis. De skal vise at de er i stand til å stille opp et problem i tillegg til å løse det. Det forventes at elevene kan etablere forbindelser innen matematikken og anvende problemløsningsmetoder hentet fra flere andre fagområder.

### *1.3.3 Situasjoner – kontekst*

Kontekst kan i matematikkens didaktikk ha to forskjellige betydninger, *oppgavekontekst* og *situasjonskontekst*. (Wedege 2000, s.3) Den første betydningen representerer virkeligheten enten i en oppgave, i et eksempel eller i annet undervisningsmateriale (ibid). Det vil i denne betydningen gjelde sammenhengen i den formulerte teksten.

Den andre betydningen omhandler kontekster for matematikkundervisning og bruken av matematikk i forskjellige situasjoner (ibid, s.4). Dette inkluderer både distansen mellom eleven og den involverte matematikk i en oppgavekontekst. Derfor vil PISA-undersøkelsen også fokusere på elevenes matematikkunnskaper for å løse oppgaver, enten det er med for eksempel skolekontekst eller med vitenskapelig kontekst. I tillegg til slike oppgaver ønsker PISA også et begrenset antall oppgaver som ikke referer til forhold utenfor matematikkens verden, "*intra-matematisk*". Selv om oppgaven *Totallsystemet* (se vedlegg) har en innledning med vitenskapelig kontekst, er den et eksempel på den type oppgaver.

> *If a task refers only to mathematical objects, symbols or structures, and makes no reference to matters outside the mathematical world, the context of the task is considered as intra-mathematical. A limited range of such tasks will be included in PISA, where the close link between the problem and the underlying mathematics is made explicit in the problem context. (OECD/PISA2001 b, s.12)*

Matematikk er et mangesidig fag. Den kan både være *ren* vitenskap og *anvendt* vitenskap inkorporert i forskjellige kontekster. I den siste rollen er matematikkens oppgave å bidra til forståelse og utvikling i "*ekstra-matematiske*" områder (Niss 1994, s.367). Den matematikken elevene til daglig møter er ikke alltid formulert i eksplisitte matematikkoppgaver, men ofte i andre situasjoner. Den underliggende matematikken ligger da i problemkonteksten som ofte kan være autentisk fra dagliglivets situasjon.

> *More typically, problems encountered in the day-to-day experience of the student are not stated in explicit mathematical terms. They refer to real world objects. These task contexts are called "extra-mathematical", and the student must translate these problem contexts into a mathematical form. Generally speaking, PISA puts an emphasis on tasks that might be encountered in some real-world situation and possess an authentic context that influences the solution and its interpretation. (OECD/PISA 2001b, s.12)*

I PISA-programmet er hensikten å måle hvor godt dagens 15-åringer er forberedt på å møte utfordringene i dagens informasjonssamfunn. Derfor skal oppgaver utviklet for PISA fylle denne funksjon.

Konteksten i en oppgave er lokalisert i en viss avstand fra eleven og inkluderer en problemstilling som krever en matematisk behandling. Hensikten for eleven blir å løse et reelt problem hentet fra dagliglivet. Når eleven går i kast med oppgaven, vil de ulike kontekster befinne seg i ulike avstander og derfor vil elevens engasjement for å løse oppgaven avhenge av den situasjon oppgaven presenteres i. I PISA-sammenheng er den nærmeste konteksten for eleven *Personlig liv* og oppgaven har da problemstilling forbundet med dette (se *Potetoppgaven* p.7.1). Deretter følger *Skoletilværelsen* etterfulgt av *Arbeidslivet* (se *Trappekonstruksjon* p.7.2). Så kommer *Nærmiljø og samfunn* (se *Barnefødsler* p.7.3) og situasjoner lengst bort fra eleven er *Vitenskapelige* kontekster. Eksempel på det er oppgaver knyttet til vitenskapelige emner som *Fibonaccitallene*.

Den oppdelingen som PISA har foretatt er sikkert i overensstemmelse med hvordan elevene i de fleste deltagende land oppfatter det. Det kan diskuteres hvorvidt 15-åringer i Norge oppfatter det på samme måte. For mange er *Arbeidslivet* en fjern framtid og flere elever vil ha større engasjement i oppgaver med kontekst hentet fra *Nærmiljø og samfunn*.

De ulike kontekstene i matematikkoppgaver utviklet for PISA kan derved plasseres i en kontinuerlig skala ut fra avstanden fra eleven avhengig av erfaringsbakgrunn. Det legges vekt på at konteksten eleven møter i den enkelte oppgave, er situasjoner en kan møte i det virkelige liv, og eleven må selv oversette problemstillingen i oppgaven til en matematisk form. Eleven må så bruke sine ferdigheter og kunnskaper i matematikk for deretter å løse oppgaven.

En interessant side ved denne avstandsplasseringen er, som tidligere nevnt, hvorvidt svarprosenten på slike oppgaver påvirkes av oppgavens kontekst. Analysen av elevbesvarelsene vil eventuelt bekrefte dette.

# 2. Matematikk i norsk skole

Ved utvidelsen av den obligatoriske skole med innføringen av ungdomsskolen, ble det besluttet at alle norske 15-åringer skulle lære matematikk. Hva slags matematikk som passet for alle 15-åringer, var den gang et sentralt spørsmål. To hovedsiktemål har preget matematikkundervisningen i etterkrigstiden. "*På den ene siden er matematikk nyttig, på den andre siden er den også et ledd i en danningsprosess*" (Gjone 1994, s.1). Denne svingningen i synet på matematikkundervisningen som en bevegelse mellom på den ene siden nytteargumentet og på den andre siden dannelsesargumentet er et resultat av hva skolepolitikere og andre har ønsket at matematikkundervisningen skal gi samfunnet.

> *Var det folkeskolens regneundervisning – med vekt på det praktiske som skulle være det dominerende element, eller var det den høgre skolens allmenndannende matematikk som skulle være grunnlaget i de øverste trinnene i den nye 9-årige skolen? (ibid, s.7)*

Jeg oppfatter ikke disse to argumentene som antagonistiske, men heller som en vektlegging. Gjennom flere forsøksprosjekt for *moderne matematikk*[3] i 1960-årene hadde forkjemperne tro på at fagets nytteverdi ble tatt vare på. Samtidig ble faget tillagt stor allmenndannende verdi, "*bl.a. med hensyn til arbeidsdisiplin og tanketrening*" (Forarbeid til Normalplan for grunnskolen s.123). I praksis viste det seg at nytteverdien ikke ble helt etter intensjonen.

> *Subjective aspects such as the students' experiences, knowledge specific to their age group, and inner representation of concepts were scarcely taken into account. (Tietze 1994, s.44)*

Det kom reaksjoner mot *moderne matematikk* også i Norge. Også her dreide det seg i første rekke om nytteverdien. Jeg arbeidet den gangen i grunnskolen og fikk et inntrykk av at elevenes regneferdigheter ofte var mangelfulle. Mange av de elevene som hadde *moderne matematikk* behersket for eksempel ikke elementær prosentregning.

I forbindelse med innføringen av mønsterplanen i 1974 ble det sterkere vektlegging på nytteverdien ved faget. Ved revisjon av mønsterplanen i 1987 ble nyttehensynet klart dominerende selv om dannelsesaspektet også ble understreket i flere sammenhenger (Gjone 1994).

Hvor står så skolen i dag? Kanskje er forsøket på å forene disse to tradisjonene en viktig intensjon med Læreplanverket for den 10-årige grunnskolen (L97). Dette synet styrkes av nettopp de nye elementene som omtales i L97. Både *det allmenndannende mennesket* som omtales i generell del av Læreplanen og *at matematikk blir et redskap elevene kan ha nytte av* i felles mål for faget understreker dette.

[3] I skolen ble den ofte omtalt som "ny matematikk"

## 2.1 Matematikk for dagens norske 15-åringer

Læreplanverket for den 10-årige grunnskolen (L97) danner grunnlaget for opplæringen bl.a. i matematikkfaget og gir rammer for undervisningen. Læreplanen er derfor det naturlige utgangspunktet når man skal sette seg inn i hvilke deler av matematikken og hvilket nivå, dagens 15-åringer møter i norske skoler. L97 skiller mellom *mål* og *målområder*.

Læreplanen har formulert seks felles mål for matematikkfaget (L97, s.158) som skal danne en basis for at elevene skal utvikle grunnleggende kunnskaper i faget. Disse målene tar blant annet sikte på at opplæringen skal legge vekt på elevenes utvikling av både holdninger, faktakunnskaper, ferdigheter, begrepsstrukturer og at elevene utvikler generelle strategier for å løse oppgaver i faget.

Det første felles mål er *at elevene utvikler et positivt forhold til matematikk, opplever faget som meningsfylt og bygger opp selvfølelse og tillit til egne muligheter i faget*. I denne sammenhengen er det viktig hvordan læreren legger opp undervisningen og derigjennom hvorledes eleven møter faget. Opplæringen har som mål at elevene skal utvikle et positivt forhold til matematikk slik at de oppfatter faget som nyttig. Det innebærer at opplæringen skal være tilrettelagt og ta utgangspunkt i den enkelte elevs erfaringer.

Det andre felles mål er *at matematikk blir et redskap elevene kan ha nytte av på skolen, i fritiden og i arbeids- og samfunnsliv*. Det allmennyttige ved matematikkfaget har ved dette fått en sentral plass i læreplanen. Elevene skal fra oppgaveeksempler oppleve faget som et redskap de selv har bruk for i dagliglivet. Betrakter man oppgaver gitt til ungdomsskoleeksamen de siste årene, ser man at andelen oppgaver med temaer fra dagliglivet har økt. (se p.2.2)

Tredje felles mål er *at elevene stimuleres til å bruke sin fantasi, sine ressurser og sine kunnskaper til å finne løsningsmetoder og alternativer gjennom undersøkende og problemløsende aktivitet og bevisste valg av verktøy og redskaper*. Elevene skal kunne legge opp en generell strategi for å løse oppgaver ved å velge både passende ferdigheter de har tilegnet seg i faget og bruke sine erfaringer. Dette vil igjen trene elevene i problemløsninger på områder som ligger utenfor matematikkfaget.

Det fjerde felles mål er *at elevene opparbeider ferdigheter i å kunne lese, formulere og formidle emner og ideer hvor det er naturlig å bruke matematikkens språk og symboler*. For at elevene skal ha nytte av matematikk, må de ha gode ferdigheter i faget enten det gjelder å tolke matematisk informasjon som blir presentert, eller formidle det utførte arbeid ved hjelp av matematikkens språk og dets symboler. De siste årenes økte andel tekstoppgaver gir elevene større utfordringer både i tekstforståelse generelt og i å trekke ut det essensielle i teksten som er av betydning for å løse oppgaven.

Det femte felles mål er *at elevene utvikler innsikt i grunnleggende begreper og metoder i matematikk, og utvikler sin evne til å se sammenhenger og strukturer og kunne forstå og bruke logiske resonnementer og trekke slutninger*. Jeg oppfatter dette som en større vektlegging på betydningen av å bygge opp begrepsstrukturer i faget. I undervisningen skal elevene utvikle sine evner til forståelse og refleksjon som viktige komponenter i det som utgjør matematisk kompetanse (se p.1.3.2). Dette bryter med tidligere antakelser om at elevene lærer ved å gjenta en regneoperasjon mange ganger.

*Den dominerende tradisjon når det gjelder undervisning i matematikk, synes å være basert på troen på at gjentatte øvelser av fakta og ferdigheter fører til bedre forståelse av begreper. (Brekke 1995, s.9)*

Eldre lærebøker og oppgavesamlinger preges av slike rutinemessige gjentagelser. I mange tilfeller var det elevenes ferdigheter i å løse oppgaver som ble bedre, mens utvikling av elevenes innsikt i grunnleggende begrepsstrukturer, ikke ble prioritert.

*Lærebøker har tradisjonelt lagt hovedvekten på eksempel-regel-metoden knyttet til fakta og ferdigheter, med øving på disse som det viktigste. Lærebøker er utformet slik blant annet med tanke på å lette arbeidet for en travelt opptatt lærer. Av samme grunn er språket gjort så enkelt som mulig. Oppgavene er svært ofte fragmentert til små isolerte steg, der målet er å øve seg slik at en kan mestre disse stegene – ett om gangen. På den måten blir elevenes aktiviteter i første rekke rettet mot dette, mens aktiviteter som retter seg mot begrepsmessige diskusjoner og refleksjoner, kommer i andre rekke. (ibid, s.18)*

L97 legger til grunn at det er flere vesentlige elementer i en læringsprosess. Tidligere kunnskaper i faget og erfaringer elevene gjør gjennom de oppgaver de blir stilt overfor, er slike elementer. I undervisningssituasjon er det elevene selv som konstruerer sine matematiske begreper, opparbeider sine ferdigheter, utvikler metoder, ser sammenhenger og trekker slutninger. Lærerens rolle er å stimulere denne prosessen.

*I skolene blir elevene undervist av et personale som har dette som sin jobb og er utdannet med dette som formål. Men læring og undervisning er ikke det samme. Læring er noe som skjer med og i eleven. Undervisning er noe som blir gjort av en annen. God undervisning setter læring i gang – men den fullbyrdes ved elevens egen innsats. Den gode lærer stimulerer denne prosessen. Elevene bygger i stor grad selv opp sine kunnskaper, opparbeider sine ferdigheter og utvikler sine holdninger. Dette arbeidet kan oppmuntres og påskyndes – eller hemmes og hindres – av andre. (L97, s.28)*

Dette er et konstruktivistisk læringssyn som baseres på at hvert enkelt individ er aktive i å konstruere sin egen virkelighet. *Når vi lærer vil vi aldri bare overta andres kunnskaper eller ferdigheter, men vi må selv være aktive konstruktører* (Sjøberg 1998 s.39). I denne prosessen er det da også viktig at elevene oppfatter undervisning som de møter i skolen, meningsfull. Å utvikle matematikkoppgaver med realistiske problemstillinger som elevene kjenner igjen fra dagliglivet, vil virke motiverende for elevene.

*Elevene konstruerer selv sine matematiske begreper. For denne begrepsdannelsen er det nødvendig å vektlegge samtale og ettertanke. Utgangspunktet bør være meningsfylte situasjoner, og oppgaver og problemer bør være realistiske slik at de virker motiverende på elevene. Elevene kan ha uferdige begreper, gjør av og til feil og viser misoppfatninger. (L.97 s.155)*

Det sjette og siste felles mål er *at elevene utvikler innsikt i matematikkens historie og fagets rolle i kultur og vitenskap*. Dette er et nytt mål i læreplanen i forhold til tidligere planer og gjenspeiler at L97 legger stor vekt på at opplæringen skal gi en god allmenndannelse. Kunnskapen om matematikkens utvikling og dens rolle som et redskap i kultur og vitenskap, gjør faget rikere for elevene. Matematikkens historie er også en rik kilde for utvikling av oppgaver som både er utfordrende og engasjerende.

Læreplanen i matematikk (L97) er en videreutvikling og fornyelse av skolenes tidligere læreplaner. En viktig del av denne fornyelsen er nettopp vektlegging av det allmennyttige i alle fag og ikke minst i matematikkfaget.

I hele grunnskolen, fra småskoletrinnet til ungdomstrinnet, er matematikk i dagliglivet et målområde (L97). Det betyr at elevene skal ha kjennskap til de grunnleggende matematiske begreper som har sammenheng med de erfaringer de gjør i hverdagen. For ungdomsskoletrinnet inkluderer dette i tillegg at elevene bruker sine matematikkunnskaper som redskap i problemoppgaver som er relatert til dagligliv og samfunnsliv.

> *Elevene skal lære å bruke sine kunnskaper i matematikk som et nyttig redskap i oppgaver og problemer i dagliglivet og samfunnslivet. Elevene skal ut fra et aktuelt tema eller problemområde kunne systematisere og formulere opplysninger med matematikkens språk, utvikle resultater ved hjelp av metoder og redskaper de rår over, og prøve dette på den aktuelle sammenhengen.(ibid, s.166)*

Matematikk som praktisk redskap for å løse problemoppgaver fra temaer i dagliglivet, er i seg selv ikke et matematisk emne på linje med algebra, geometri osv. Men her får elevene stor bruk for den matematiske kompetanse de tilegner seg i undervisningen. Elevene vil her se hvilken nytte de har av å kunne mestre faget i situasjoner som ikke er direkte relatert til skolesituasjonen. Med utgangspunkt i et tema fra dagliglivet skal elevene systematisere de informasjoner de henter fra for eksempel en tekst for så å bruke matematikk som et redskap for å løse den aktuelle oppgaven.

Ved siden av *Matematikk i dagliglivet* er det nedfelt fire andre målområder for matematikkfaget på ungdomstrinnet (ibid): *Tall og algebra*, *Geometri*, *Behandling av data* og *Grafer og funksjoner*.

*Tall og algebra*. Elevene stifter her bekjentskap med algebra ved å bruke og behandle enkle bokstavuttrykk. Bruken av bokstavuttrykk blir en videreføring av elevenes tidligere kjennskap til tallregning, men her vil de også bruke bokstavene som symboler for ukjente størrelser i forbindelse med formler og likninger. I siste klasse på ungdomstrinnet (ibid, s.169,170) vil også bokstavregningen inkludere bruken av variable størrelser for å formulere og bevise generelle sammenhenger. De skal også stille opp og tolke likninger, både med én og to ukjente, kontrollere løsningen og vurdere resultatet. Elevenes møte med tall og algebra i kulturell og historisk sammenheng, gir elevene muligheter til å møte tall med spesielle egenskaper (se oppgaven *Fibonaccitallene* p.7.5), og på den måten knytte matematikk til andre fag.

*Geometri*. I L97 er dette sterkere vektlagt for 15-åringer enn i den foregående planen selv om geometrien fortsatt ikke har den posisjon den hadde for 30 år siden. Dessuten fremstår emnet noe annerledes enn i tidligere læreplaner, spesielt de utforskende sidene ved geometri. Elevene skal arbeide med figurer og former og deres egenskaper. Det siste året på ungdomsskolen skal elevene *utføre og beskrive geometriske avbildninger, slik som parallellforskyvning, speiling og rotasjon og utnytte geometriske avbildninger til å skape og analysere mønstre*.(L97, s.170)

*Behandling av data*. Ved siden av å kunne bruke forskjellige dataprogrammer, database og regneark, må elevene også kunne hente inn, tolke, vurdere og presentere informasjoner hentet fra data. Ut fra statistiske undersøkelser presentert i tabellform og annet datamateriell skal elevene både kunne lage diagrammer og presentere informasjon ved hjelp av enkle statistiske beregninger. De skal også arbeide med begreper innen sannsynlighet.

*Grafer og funksjoner*. Dette er et emne elevene stifter bekjentskap med når de begynner på ungdomsskolen. Bruken av grafer og funksjoner vil ved siden av i rent matematisk sammenheng også benyttes til å undersøke og beskrive situasjoner og sammenhenger ved praktiske problemer i dagliglivet. 15-åringene skal også utnytte funksjonsbegrepet til å løse likninger og ulikheter grafisk. Dessuten vil de arbeide med oppgaver der proporsjonalitet og omvendt proporsjonalitet med lineære og kvadratiske sammenhenger forekommer.

Matematikkoppgavene som er utprøvd i forbindelse med denne hovedoppgaven ligger innenfor læreplanverket for den 10-årige grunnskolen og representerer alle tre kompetanseklassene. (se p.1.3.2). Noen oppgaver fra kompetanseklasse 3 har lav prosent riktige svar (høy vanskegrad), noe som tyder på at oppgavene var vanskelig for de fleste elevene selv om de ligger innenfor læreplanen. Selv om ungdomstrinnet har det største omfang av sentralt fastsatt lærestoff i grunnskolen, gis det rom for å arbeide i ulik dybde for å få en individuell tilpasning. Det vil derved være naturlig at det utvikles egne oppgaver for den gruppen elever som ønsker å arbeide med mer "utfordrende" oppgaver. Ansvaret for å tilby elevene slike oppgaver ligger fortsatt hos den enkelte lærer.

## 2.2 Oppgaver gitt til ungdomsskoleeksamen i forhold til PISA-oppgaver

Med *Reform 94* for den videregående skolen og *Læreplanverket for den 10-årige grunnskolen* (L97) gjennomgikk norsk skole store endringer på 90-tallet. I grunnskolen begynte de første årskullene å følge L97 skoleåret 1997/1998. I løpet av tre år var planen innført på alle klassetrinn.

Reformen med utgangspunkt i L97 og gjeldende forskrift og regelverk medførte også store endringer i innholdet og formen på den skriftlige avgangsprøven i matematikk for grunnskolen. Avgangsprøven blir oppfattet som et viktig ledd i kvalitetssikringen av opplæringen både hva gjelder innhold og arbeidsmåte.

> *Det føregår kontinuerleg utviklingsarbeid for å finna fram til nye prøveformer og oppgåver som best mogeleg avspeglar målsetjinga i gjeldande læreplanar for faga. Det skjer også ei stadig utvikling når det gjeld bruk av hjelpemiddel ved prøvene. (Statens utdanningskontor i Hordaland)*

Både i 1998 og 1999 hadde eksamen "tradisjonell" form der elevene fikk anledning til begrenset bruk av lommeregner. Prøven besto av to oppgavehefter og etter innlevering av første heftet kunne elevene fritt benytte lommeregner for siste heftet.

I 2000 ser vi at reformen har kommet med full virkning. Avgangsprøven fra og med dette året er endret for bedre å kunne måle den kompetansen elevene skal ha etter L97. Prøven består nå av tre oppgavehefter, og i tillegg får elevene utdelt en informasjonsbrosjyre en og en halv dag før prøvedagen som de også skal ta med seg til avgangsprøven. Der får elevene de nødvendige opplysningene for å løse enkelte av oppgavene. På den måten skal elevene vise sin kompetanse i matematikk som et nyttig redskap i "problemer i dagligliv og samfunnsliv". Innholdet i informasjonsbrosjyren i 2000 spente fra geografiske opplysninger som kart over nasjonalparker, prisutviklingen over viktige matvarer de siste hundre årene, priser for bruk av mobiltelefon til andre opplysninger av ulike slag. På prøven ble det også gitt oppgaver der elevene fikk vise sin kreativitet ved at de selv skulle lage oppgavetekst ut fra gitte opplysninger, enten det var oppgitte tall, tekst formulert som en salgsannonse, eller figurer. Disse oppgavene skulle så også løses. På den måten fikk man grunnlag for å vurdere elevenes evner til å gjøre bruk av et matematisk språk. Dette er en oppgavetype som i liten grad blir gitt i matematikkdelen for PISA-undersøkelsen. En grunn til det kan være at slike oppgaver gir problemer i retteprosedyren. Reliabiliteten er viktig å vareta spesielt i større internasjonale undersøkelser.

En annen viktig reform med avgangsprøven i 2000 er at bruk av egenprodusert elevbok[4] under eksamen nå er tillatt. Elevene skal selv samle stoff gjennom året, både i form av egne notater, regler eller som eksempler på anvendelse. Betenkelighetene ved denne delen av reformen kan være flere. En helhetskompetanse i faget inkluderer også elevens forarbeid, planlegging og strategisk valg av emner i en egenprodusert elevbok, men den form elevboka kan få varierer fra skole til skole avhengig av lærerens instruksjoner om omfang og innhold. Det er en fare for at mange lærere er usikre på hvilket omfang en slik elevbok bør ha. Som følge av bl.a. dette, kan resultatet for eksempel bli at noen elever fyller opp sine elevbøker med mange eksempler på hvordan en oppgave skal løses. For eleven blir eksamen da i realiteten å lete frem en mønsterløsning som passer til oppgaven. Derved er elevboken en oppskriftsbok på hvordan en oppgave skal løses og man får ikke vurdert om eleven nødvendigvis har full forståelse av den matematikk som anvendes i oppgaveløsingen.

Lærebokforfatterne av matematikkbøkene utgitt på Aschehoug forlag har fått flere henvendelser angående bruk av elevbok til avgangsprøven.

> *Vi har hatt kontakt med en del matematikklærere. Det kan se ut til at det råder en skepsis blant flere av lærerne til bruk av regelbok. Elevene har arbeidet med regelbok i ungdomsskolen, og erfaringene kan tyde på at dette har ført til at en del elever har lagt seg til en læringsstrategi der det å "kunne noe", er blitt det samme som å finne et eksempel i regelboka og bruke det som mal. Tankeprosessen som ligger foran oppstillingen av et regnestykke, ser ut til å mangle oftere enn før. Elevene ser ikke ut til å forstå hva vi spør etter når vi ber dem forklare. En løsning på dette problemet kan være å veksle mellom regelboksprøver og regelboksfrie prøver, eller å la elevene begrense "regelboka" til f. eks. en A4-side på enkelte prøver. (Erstad & Bjørnsgård 2000, s.16)*

Bruken av egenprodusert elevbok ved avgangsprøven kan altså forårsake forskjeller avhengig av det arbeid som eleven har nedlagt på forhånd. På tross av det kan det argumenteres for betydningen av elevbok. Elevene får under avgangsprøven nyttiggjøre seg de hjelpemidler som de normalt har tilgang til i dagliglivet.

Til forskjell fra den skriftlige avgangsprøven til grunnskolen, vil elever som er med på PISA-undersøkelsen ikke ha anledning til å bruke egenprodusert elevbok på testen. Derved er det grunn til å anta at med samme type oppgaver vil prestasjonen en elev yter på en slik undersøkelse være forskjellig fra det som ytes på den norske avgangsprøven.

I motsetning til PISA-undersøkelsen, består avgangsprøven av alternative oppgaver med ulik poengsetting, og til hver oppgave er det oppgitt hvor mange poeng eleven vil oppnå dersom de respektive oppgaver løses. Elever med et høyere ambisjonsnivå vil kunne konsentrere seg om og velge oppgavene med flest poeng. For noen kan det bli et sjansespill. Andre elever som ikke føler seg helt trygge på faget, kan antas å velge alternative oppgaver med færre poeng. I tillegg skal de velge ut noen få oppgaver i det siste heftet, og de poeng de kan oppnå vil da også avhenge av hvilke oppgaver de velger å løse. Poengsettingen vil også her styre det valget noen gjør. Derved kan en elev med manglende faglig selvtillit bli skremt fra å prøve seg på en oppgave han eller hun faktisk normalt ville klare hvis ingen "fortalte" at oppgaven er vanskelig, mens en elev med en noe overdreven selvtillit, vil bruke

[4] I diskusjoner om bruk av elevbok til eksamen, blir den ofte omtalt som regelbok

uforholdsmessig mye tid på oppgaver som er i overkant av elevens kompetanseområde, og derved ikke få tid til i stedet å løse flere enklere oppgaver.

Det er ikke tradisjon for å benytte flervalgsoppgaver til eksamen i matematikk i Norge. Ved avgangsprøven for 2001 dukket det imidlertid opp en flervalgsoppgave som en del av en oppgaveenhet (Læringssenteret 2001b). I denne oppgaven skulle elevene krysse av for det funksjonsutrykket som passet til en rett linje tegnet i et koordinatsystem. Oppgaven er ikke ulik den elevene fikk presentert i pilottesten noen uker før avgangsprøven (se spørsmål 4 i p.7.1).

Den hyppige bruk av flervalgsoppgaver i internasjonale tester vil forhåpentlig føre til at vi i de nærmeste årene vil se flere oppgaver av dette formatet. Fordeler og ulemper ved slike oppgaver blir nærmere gjennomgått i p.3.2 og p.3.4.

En sammenligning av mine oppgaveforslag som ble testet på noen grupper avgangselever i ungdomsskolen og de skriftlige avgangsprøvene gitt fra årene 2000 til 2003, viser store likheter. Det er oppgavenes formål som er forskjellig. Den skriftlige avgangsprøvens formål er å rangere den enkelte elev i en endimensjonal skala for å fastsette en karakter. Den tar sikte på å måle elevenes kompetanse bredest mulig i forhold til fellesmål og hovedmål i fagplanen i L97. Ved sensureringen av besvarelsen tas det ikke bare hensyn til poengsummen den enkelte elev oppnår på prøven, men også helhetsinntrykket tillegges stor vekt ved fastsetting av karakter. I henhold til *Sensorveiledning for skriftlig avgangsprøve i matematikk 2002* skal en ved helhetsinntrykk vektlegge blant annet:

- *Hvilke oppgaver eleven har løst (vanskegrad)*
- *Hvordan eleven mestret de ulike områder av faget*
- *Hvilke metoder/framgangsmåter eleven har brukt*
- *Elevens evne til å resonnere logisk*
- *Kreativitet*
- *Om løsningen av oppgavene er fremstilt på en oversiktlig og klar måte*
- *Hvordan besvarelsen er ført*

*(Læringssenteret 2001a, s.2)*

PISA-undersøkelsen går ut på å sammenligne 15-åringers, altså de samme elevgruppers, kompetanse i matematikk både mellom land som er med i prosjektet, og mellom jenter og gutter i samme land. Intensjonen med pilottesten var at oppgaveforslagene, og ikke den enkelte elevs prestasjon, skulle testes. I den grad enkeltelevers prestasjoner er av interesse, er det først og fremst i forbindelse med analysering av oppgavene for å danne grunnlag for senere å bearbeide disse. I likhet med de generelle prøver elevene møter gjennom skolegangen, er det også her viktig å undersøke om de elever som svarer riktig på en enkeltoppgave, også har en høyere totalskår på hele testen enn de som ikke svarer riktig på enkeltoppgaven (se 4.1.5). Dette er viktig å analysere for å kunne vurdere om oppgavene har en bra diskrimineringsevne (se 4.1.4), hvis hensikten med testen er å måle elevgruppers kompetanse i matematikk. Vi vil senere i analysen se at ikke alle mine oppgaveforslag svarer til forventningen og vil av den grunn være lite velegnet til en kunnskapstest. Slike oppgaver kan ha andre akseptable funksjoner som f. eks diagnostiske tester.

En annen viktig forskjell, er at de oppgaveforslagene som er utviklet for PISA, ikke nødvendigvis tar utgangspunkt i fagplanen i L97 siden de ulike land har forskjellige fagplaner og ”pensum”, mens den skriftlige avgangsprøven i Norge naturlig nok består av oppgaver som representerer de respektive sider ved fagplanen. Det må legges til at det er mange likhetstrekk mellom de testoppgaver som presenteres i PISA-undersøkelsen og de eksamensoppgaver som er gitt etter L97. Dette er et resultat av en utvikling gjennom år der tendensen er at matematikkoppgaver etter L97 får stadig flere trekk til felles med andre lands oppgaver selv om flere land er opptatt av ”rene” oppgaver. Denne utviklingen er kanskje et resultat av flere internasjonale undersøkelsene vi har hatt de siste årene?

Et viktig prinsipp for de oppgaver som utvikles til PISA-undersøkelsen, er at muligheten for å besvare de spørsmål som presenteres i en oppgave, ikke bør være avhengig av om eleven mestret tidligere spørsmål i oppgaven. Dette gjelder ikke for avgangsprøven selv om man ser en tendens til en utvikling mot dette også her. Ved avgangsprøven i 1999, den siste før reformen var gjennomført, var det 4 spørsmål som forutsatte at eleven mestret tidligere spørsmål i oppgaven. Antallet var redusert til 1 for prøven i 2002.

De oppgaver som er testet og er med i denne hovedoppgaven, ligger alle innenfor fagplanen i L97, men dekker langt fra alle fellesmål og hovedmomenter. Det er bl.a. ikke tatt med geometriske konstruksjoner og tilhørende beregninger av sider og vinkler. Oppgavene er utarbeidet på grunnlag av de oppgavetyper en ønsket å utvikle etter rammeverk for PISA-undersøkelsen. Hvordan noen enkelte oppgaver vil fungere i en skriftlig avgangsprøve blir vurdert i kapittel 7 (presentasjon og diskusjon av oppgavene).

# 3. Oppgaveformat

Spørsmålene i denne pilottesten er organisert i oppgaveenheter og er stort sett knyttet til en innledende tekst, enten fra situasjoner i dagliglivet, problemstillinger hentet fra utsnitt fra avis, forskrifter eller matematiske temaer. Siden oppgavespørsmålene tar utgangspunkt i den innledende teksten, bør elevene lese teksten grundig før de svarer på spørsmålene. Spørsmålene representerer alle de tre kompetanseklasser (se p.1.3.2). De varierer fra de enkleste problemstillinger til spørsmål med større grad av kompleksitet, men det er i de fleste oppgavene lagt vekt på at spørsmålet skal være slik at en ikke er avhengig av svar fra et tidligere spørsmål i oppgaven for å kunne gå videre i oppgaven.

Det er benyttet to hovedkategorier av oppgaveformat, flervalgsoppgaver og åpne oppgaver. I de tradisjonelt åpne oppgaver som er mye brukt i Norge, må eleven selv foreslå et svar, enten bare et tall (kortsvar) eller eventuelt med større utregninger, forklaringer, bevis eller begrunnelser (langsvar). Eleven har derfor stor frihet til å velge både form og innhold i besvarelsen.

Av flervalgsoppgaver er det igjen to typer. I den enkleste skal elevene velge det riktige svaret blant flere alternativer som alle virker sannsynlige, men bare én er riktig. At alle alternativene umiddelbart kan virke riktige, er viktig for å forhindre at eleven skal kunne velge riktig svar ved å eliminere urimelige alternativer (de gale alternativene kalles distraktorer). I den andre typen, de sammensatte flervalgsoppgavene, må eleven svare **riktig/galt** eller **ja/nei** på en serie utsagn eller spørsmål som er knyttet til de opplysninger eleven finner i den innledende teksten til oppgaveenheten.

## 3.1 Oppgaveformat ved diagnostisering

Formålet med å konstruere diagnostiske oppgaver er å studere elevenes enten intuitive oppfatninger eller manglende kunnskaper i matematikk, og ikke å rangere elevene etter et poengsystem. Det er derfor viktig å gjøre elevene oppmerksom på forskjellen mellom en tradisjonell prøve og en diagnostisk test, og hva som er hensikten med en slik test. Tolkningen av besvarelsen ved slike tester skal gi grunnlag for å dra slutninger om elevens tenkning rundt de problemer som oppgaven presenterer. Da er det viktig å forstå forskjellen på feil elever gjør på en oppgave, og de misoppfatninger de har.

> *En feil kan komme mer eller mindre tilfeldig, fordi en ikke er oppmerksom nok eller ikke leser oppgaven godt nok osv. Misoppfatninger er ikke tilfeldige. Bak dem ligger det en bestemt tenkning – en ide – som en bruker nokså konsekvent.*
> *(Brekke 1995, s.11)*

Mange elever har sine private forestillinger i matematikk som tilsynelatende kan ha fungert utmerket i mange sammenheng. Forestillingene har utviklet seg over tid og kan i flere situasjoner tilsynelatende gi forklaringer på matematiske problemer selv om de ikke er riktige. Slike forestillinger er et resultat av elevenes egen tenkning basert på erfaringer de har gjort, og oppleves som riktige fordi de har gitt elevene en forklaring på et problem. Disse forestillingene er et resultat av ufullstendige tanker knyttet til begrepsstrukturer og blir i matematikkfaget kalt misoppfatninger.

En diagnostisk oppgave i flervalgsformat bør inneholde distraktorer basert på slike misoppfatninger. Hensikten er å fremheve disse misoppfatningene for på det vis å kartlegge elevenes ståsted kunnskapsmessig. Selv om flervalgsformatet brukes, bør elevene også i noen av oppgavene vise hvordan de kom fram til svaret. Utregninger og forklaringer gir verdifull informasjon om hvordan elevene tenkte. Testens funksjon blir da å se om noen av distraktorene velges hyppig. Dette er i så fall en indikasjon på at det er alternative forestillinger (misoppfatninger) blant elevene.

## 3.2 Flervalgsoppgaver i matematikk

Et viktig spørsmål ved utvikling av oppgaveenheter er hvilken funksjon oppgavene skal ha. Dersom oppgavens funksjon er å diagnostisere elevenes matematikkunnskap, vil ofte en flervalgsoppgave foretrekkes. Ved å følge svarmønsteret på flere etterfølgende oppgaver, vil eventuelle misoppfatninger komme frem. I de tradisjonelt åpne oppgavene der elevene selv formulerer sine svar, vil flere usikkerhetsmoment komme inn. Både formuleringsmåte og evnen til å formidle, er forskjellige blant elevene, og det er ikke sikkert at sensors tolking av svaret er i samsvar med hvordan eleven tenkte. Den som retter oppgaven vil tolke ut fra det vedkommende antar må være elevens egen tenkning. Det er derfor fare for at en besvarelse blir gjenstand for forskjellige tolkninger. Det kan derfor bli vanskelig å få tak på hva eleven egentlig mente.

Mens man i Norge ikke har tradisjon for flervalgsoppgaver, brukes dette i mange land bl.a. i forbindelse med eksamen og i internasjonale tester. Da er det løsningsfrekvens, hvor mange og hvem som svarer riktig, som er av interesse. Gjennom internasjonale undersøkelser som PISA og TIMSS[5] (The Third International Mathematics and Science Study) får norske elever stifte bekjentskap med flervalgsoppgaver.

Det er flere fordeler med denne type prøver. Her kan nevnes:

- Retteprosedyren forenkles, og sjansen for rettefeil minimeres, noe som igjen øker sensorreliabiliteten.
- Det letter arbeidet ved den statistiske behandlingen.
- Det er mulig å dekke en større del av pensum siden eleven sparer tid på besvarelsen.
- Ved en diagnostisk undersøkelse vil det være mulig å legge inn antatte misoppfatninger i de ikke riktige alternativene, kalt distraktorer. Med flere påfølgende flervalgsoppgaver kan man sirkle rundt disse misoppfatninger, og ved statistisk behandling med bruk av krysstabeller, er det enklere å følge elevens svarmønster. Men det er krevende å konstruere gode flervalgsoppgaver.

[5] Undersøkelsen i 2003 har navnet Trends in International Mathematics and Science Study

Andre fordeler med flervalgsoppgaver er at svaralternativene kan hjelpe elevene til å forstå spørsmålets intensjon (Olsen, Turmo, & Lie 2001, s.408). Dette gjelder spesielt når oppgavens innledning framstår som vanskelig.

> *An interesting hypothesis as that MC[6] items, to some extent, reflect a Vygotskian perspective of knowledge. If a question is stated and the student is left alone to answer this question in his or her own words there is a possibility that some misunderstanding will occur. By giving response alternatives, the student is provided with additional information on how to interpret the question. In this sense, the distractors have the same function as a conversation partner, hindering some of the possible misinterpretations. (ibid)*

Argumentene mot bruk av flervalgsoppgaver er mange. Noen elever vil tippe svar uten å kontrollere ved egne utregninger. Skulle eleven tippe riktig, er det jo selvfølgelig ingen garanti for at eleven har mestret oppgaven. I motsetning til den enkle varianten av flervalgsoppgaver, vil det være vanskeligere å tippe riktig i de sammensatte flervalgsoppgavene. Dersom det er 5 svaralternativ i en flervalgsoppgave av den enkle varianten, har eleven 1/5 sannsynlighet for å tippe riktig, mens i den sammensatte varianten vil sannsynligheten reduseres betraktelig. Med de tradisjonelt åpne oppgavene (med langsvar) vil man i besvarelsen følge elevens tenkning og selv med et feil svar vil man kunne se om eleven behersker den matematikken som ligger i oppgaven. Feilen kan jo bero på en ubetydelig regnefeil. En annen svakhet ved flervalgsoppgaver er at elevene ved slike oppgaver vanskelig kan testes i evnen til å uttrykke seg faglig.

## 3.3 Bruk av distraktorer

I den enkle varianten av flervalgsoppgaver der eleven krysser av det korrekte svaret blant en rekke forslag til svar, har de gale forslagene den hensikt å distrahere elever som er usikre på hvilket svar som er riktig. De gale forslagene kalles distraktorer og er ikke tilfeldig valgt.

Ved den kvantitative analysen av pilotoppgavene for TIMSS ble blant annet følgende kriterier lagt til grunn.

- *Antall riktige svar på en oppgave skal ligge over gjettenivå.*
- *En distraktor bør ikke ha høyere svarfrekvens enn det riktige alternativet.*
- *Distraktorene skal imidlertid være så "gode" at svarfrekvensen på hver av dem bør være minst 5 %.*
- *"Point-biserial" korrelasjonen[7] for det riktige svaralternativet skal være høyere enn 0,2.*
- *"Point-biserial" korrelasjonen for distraktorene skal ikke overstige 0,00.*

*(Kleve 1994, s.9)*

---

[6] Flervalgsoppgave. Min tilføyelse

[7] Se p.4.1.4 Min tilføyelse

Bruken av distraktorer i de enkle flervalgsoppgavene kan ha forskjellige funksjoner. Ved valg av distraktorer er prøvens intensjon viktig. Hvis hensikt er å måle hvor riktig eleven svarer, velger man distraktorer som eleven ikke lett eliminerer, men som kan virke tilsynelatende riktig.

Valg av gode distraktorer er vanskelig ved utvikling av flervalgsoppgaver.

> *When preparing a multiple-choice test item, making distractors plausible is an important consideration. If distractors are implausible, they will not serve a useful function in measurement, but will usually be easily avoided by examinees. (Osterlind 1992, s.158)*

Det kan være et diagnostisk element i alle flervalgsoppgaver, men i en diagnostisk flervalgsoppgave har distraktorene en dypere betydning. Her vil problemet være å vurdere i hvilken grad de ulike distraktorene påvirker svaret, og hvilke konklusjoner som kan trekkes av dette. Denne type oppgaver krever mye planlegging med grundig forarbeid og inngående kjennskap til elevenes misoppfatninger. Disse misoppfatningene kan fungere som utmerkede distraktorer. Her kan læreren som konstruerer flervalgsoppgaven bruke både misoppfatninger som identifiseres i undervisningen og misoppfatninger som det er allment kjent at oppstår blant den aktuelle elevgruppen.

## 3.4 Kombinerte oppgaver i matematikk

I en oppgaveenhet blir flere sider ved et tema belyst. Ved bruk av flervalgsoppgaver er det mulig å dekke et større område av temaet hvis eleven har begrenset tid på besvarelsen. Man kan kombinere bruk av oppgaver i flervalgsformat og åpne format, der flervalgsformatet tester de lavere kategorier i Blooms kognitive taksonomi[8]. Det er vanlig at flervalgs-oppgavene inneholder tester av faktakunnskaper og ferdigheter. Mer problematisk blir det å utvikle flervalgsoppgaver som tester elevens begrepsoppfatninger der eleven får vise om han/hun forstår og kan anvende sine kunnskaper presentert i en besvarelse. Dette gjelder spesielt der eleven skal vise sin evne til å kombinere, relatere, presisere, generalisere og til slutt, trekke slutninger. I de tradisjonelt åpne oppgavene er det enklere å teste slike ”kreative” prosesser.

> *Et kjernepunkt i norsk tradisjon er jo nettopp at vi vil ha oppgaver som tester ulike aspekter ved faget, og ikke minst elevenes evne til selvstendig å kunne presentere resonnementer (Angell og Lie 1993, s.8).*

Ved å kombinere flervalgsoppgaver med åpne oppgaver er det lettere å utvikle oppgaveenheter som også går mer i dybden av det aktuelle temaet.

Det er mulig å kombinere oppgaveformatene i en oppgaveenhet. Flere av mine oppgave-forslag til PISA-undersøkelsen inneholder slike kombinasjoner. En annen kombinasjons-mulighet er at eleven må begrunne sitt valg av svaralternativ i en flervalgsoppgave og på den måten føre en argumentasjon. Det kan også i samme oppgave kreves en begrunnelse for hvorfor eleven ikke valgte distraktorene. Da faller imidlertid mye av argumentet for at

[8] Blooms taksonomi er et ”trappetrinnsystem” der de høyere nivåer bygger på de lavere, og elevene blir vurdert i forhold til hvor høyt i systemet de er kommet. Nivåene er kunnskap, forståelse, anvendelse, analyse, syntese og vurdering, der vurdering er det høyeste nivået

flervalgsoppgaver er tidsbesparende bort, siden det er mer tidkrevende å løse slike oppgaver enn vanlige flervalgsoppgaver.

Å lage gode diagnostiske tester ved å bruke åpne oppgaver forutsetter et grundig forarbeid, og koding av svaralternativene for diagnostisering etter testen er ikke mindre krevende. Ved å benytte flervalgsformat vil en del usikkerhetsmomenter ved rettingen unngås selv om tolkningen av svarene i en diagnostisk test, kan by på utfordringer.

# 4. Metodeteori

Noen sentrale statistiske begreper jeg benytter vil her bli belyst, og i den sammenheng vil jeg i dette kapitlet gjennomgå de prosedyrer som ligger til grunn for målinger og vurderinger av elevenes oppgaveløsninger. Presentasjonen vil i stor grad støtte seg på tradisjonell statistikk, spesielt den deskriptive statistikken. Det sentrale i testteorien er å klargjøre metoder forbundet med slike målinger, enten det gjelder enkeltspørsmål eller oppgavesett i sin helhet. I internasjonale undersøkelser som for eksempel PISA, er en av intensjonene å sammenligne elevprestasjonene fra de deltagende landene. De statistiske metoder jeg vil bruke på min analyse, vil ikke nødvendigvis være de samme som blir brukt i PISA. Dette er en følge av de problemstillinger jeg har satt i min hovedoppgave. Eksempel på det er å undersøke om noen av oppgavene som ikke kom med til Generalprøven, kan ha andre funksjoner. Ved å følge svarmønstret mellom spørsmålene, vil analysen også avdekke andre sider ved oppgavene. Til denne analysen vil jeg i stor grad benytte krysstabeller etterfulgt av Kji-kvadrattest.

Min intensjon var i utgangspunktet å utvikle oppgaver som primært hadde til hensikt å diskriminere elevprestasjoner som en metode for å gradere elevene i faget. Et annet viktig felt er å utvikle oppgaver som har til hensikt å fungere som diagnostiske instrument for på den måten å kartlegge misoppfatninger enten hos enkeltelever eller grupper av elever. Dette er et område innen matematikkdidaktikken med et stort utviklingspotensial.

Jeg vil også undersøke om mine oppgaveforslag fungerer bra psykometrisk. Denne metoden har sin opprinnelse fra læren om psykologiske målinger, og refererer til statistiske metoder i forbindelse med test. Det er et abstrakt psykologisk begrep, *construct*, som er den variabel vi ønsker å måle. Dette kan i matematikkdidaktikken gjelde dyktighet på forskjellige områder av matematikken. Eksempel på det er *Fibonaccitallene* (se p.7.5). Ut fra en innledende tekst, fulgt av noen figurer og spørsmål, skal elevene selv frambringe en tallfølge. Andre eksempler på slike construct er *evne til matematisk problemløsing*. (Angell 1996, s.76)

## 4.1 Testteori

Statistisk analyse går bl.a. ut på å undersøke om de observerte forskjeller er større enn de som kan oppstå ved tilfeldige variasjoner. I de situasjoner der forskjellen er påfallende er det også viktig å undersøke hvor signifikante slike forskjeller er. Det sier seg selv at reliabiliteten (se p.4.2 og kapittel 8) eller påliteligheten i slike undersøkelser er viktig. Det er jo de virkelige forskjeller som er av interesse, og ikke de som tilsynelatende kommer fram. Skal vi måle det vi har til hensikt å måle, må påliteligheten og nøyaktigheten være høy. Det vil da si at både de tilfeldige feil må være små og de systematiske feil minst mulig.

Dette kan uttrykkes i en enkel likning:

$$M_o = M_r + M_s + M_t$$

der $M_0$ = observert verdi

$M_r$ = "sann" verdi

$M_s$ = systematiske feil

$M_t$ = tilfeldige feil

Selv om $M_0$ er den observerte verdien vi får i en test, er det $M_r$ vi ønsker å måle. Med maksimal validitet (se p.4.3) vil $M_s = 0$ og hvis målingen er reliabel, vil tilfeldige feil være nær null. Men i praksis vil dette være nesten umulig, så vi kan her snakke om at det er grader av både reliabilitet og validitet i forbindelse med test.

Tilfeldige feil kan enten inntreffe under testen eller etter gjennomføringen, ved sensureringen. Noen elevgrupper kan bli distrahert av forstyrrelser under prøven, eller de kan ha manglende motivasjon for den type prøver. Dette kan ha negativ innvirkning på resultatet. Tilfeldige feil kan innvirke både positivt og negativt på den observerte verdien, og vil influere på testens reliabilitet.

> *Poenget med slike feil er at hvis en kunne tenke seg at eksaminanden kunne ta den samme eksamen på et annet tidspunkt, ville ikke de samme tilfeldige feilene bli repetert, selv om nye tilfeldige feil ville oppstå. (Angell 1996, s.83)*

Rette- eller sensureringsfeil etter testens gjennomføring er eksempel på tilfeldige feil eleven ikke har innflytelse over. Det er større farer for sensureringsfeil på åpne oppgaver enn på flervalgsoppgaver med forenklet retteprosedyre. Det er viktig at man er klar over dette i retteprosessen. Samtidig kan en flervalgsoppgave i større grad appellere til gjetting. Hvis en elev velger å gjette på alternativene i en flervalgsoppgave framfor å bruke sine kunnskaper, er dette også et eksempel på tilfeldige feil. Siden det er den observerte verdi man forholder seg til i en test, er det ønskelig med tester der tilfeldige feil blir minimert.

Minoritetsspråklige elever kan ha større problemer med å sette seg inn i oppgaver med mye tekst enn andre elever, noe som vil påvirke prestasjonen. Siden det i noen elevgrupper er stor andel elever med minoritetsspråklig bakgrunn, vil en test av disse i større grad måle elevenes evne til tekstforståelse, mens viktige sider ved elevenes faglige kompetanse i matematikk kan tapes. På den måten kan vi bli lurt til å finne tilsynelatende kunnskapsforskjeller som ikke nødvendigvis eksisterer. En annen oppgaveformulering med enklere tekst på samme tema ville kanskje føre til andre resultater. I denne hovedoppgaven er det ikke tatt hensyn til slike mulige språklige komplikasjoner siden elevenes språklige bakgrunn ikke er registrert.

En test som måler noe annet enn det som var intensjonen, er eksempel på systematiske feil og vil influere på testens validitet. Den praktiske nytteverdien av slike tester blir da også redusert.

### 4.1.1 Oppgavens vanskegrad

En oppgaves vanskegrad er det samme som andelen som har svart riktig på oppgaven, og uttrykkes med en p-verdi (må ikke forveksles med sannsynligheten for å få et resultat gitt at nullhypotesen er sann) mellom 0 og 1. Oppgavesvaret er kodet med 1 for riktig besvarelse og 0 for ikke riktig besvarelse. Vi går fram på samme måte som i p.4.1.2 ved å kode svarene dikotomt. Vi får derved fram en gjennomsnittsverdi eller en andel for de som har svart riktig på oppgaven. Dette omtales ofte som løsningsfrekvens. En ”lett” oppgave vil da få en høyere p-verdi, mens en vanskelig oppgave vil få en lavere p-verdi, noe som umiddelbart kanskje kan virke noe forvirrende. Hvis vi setter $p_1$ som vanskegrad, dvs. andelen som har svart riktig, og $p_0$ som andelen som ikke har svart riktig, vil oppgavevariansen bli:

$$\sigma^2 = p_1(1 - p_1) = p_1 p_0$$

Denne variansen gjør vi bruk av når vi vil se prestasjonen på en oppgave i relasjon til prestasjonen på hele testen. På denne måten vil vi finne de oppgaver som diskriminerer best (se ”Point-biserial” korrelasjon p.4.1.4 ).

Det må også legges til at graden av pålitelighet vi har ved å regne ut en oppgaves vanskegrad selvfølgelig er avhengig av antall testpersoner i undersøkelsen.

### *4.1.2 Analyse av én variabel*

Ved en oppgaveundersøkelse der svaret på oppgaven enten er riktig eller galt, vil $p$ være andelen riktige svar, og vil også være sannsynligheten for riktig besvarelse på oppgaveenheten i undersøkelsen.

Vi lar

$x$ = 1 for riktig besvarelse

$x$ = 0 for galt svar

$n$ = antall besvarelser

Da vil

$\sum x$ = antall riktige besvarelser

Med andre ord har vi gjennomsnittet av besvarelsen gitt ved:

$$\bar{x} = \frac{\sum x}{n} = \hat{p} \qquad \text{der } \hat{p} \text{ er et punktestimat for } p$$

Det kan være en svakhet ved en undersøkelse dersom $p$ estimeres på grunnlag av et begrenset antall prøvekandidater. Det er da viktig at man tilstreber et tilfeldig utvalg av elever. Jeg har så langt som mulig tatt hensyn til det ved å gjennomføre testen på tre skoler med til sammen 143 prøvekandidater. Sentralgrenseteoremet sier at lar vi $n$ være et stort antall uavhengige observasjoner eller målinger med samme vilkårlige sannsynlighetsfordeling, vil gjennomsnittet av målingene være tilnærmet normalfordelt. Dette sentralgrenseteoremet gjør dermed at normalfordelingen er så viktig i statistikken. Ved et stort utvalg prøvekandidater med de samme forutsetninger (som samme læreplan og vektlegging av de samme berørte områdene i faget), vil prestasjonen for grupper av elever være tilnærmet normalfordelt. Vi kan bruke normalfordelingstabellen til å finne sannsynligheten for at gjennomsnittet ligger innenfor et begrenset intervall. Dette kan vi fastslå fordi summen av mange tilfeldigheter er normalfordelt.

Standardavviket til estimatoren $\hat{p}$ kalles også for standardfeilen til $\hat{p}$ og vil ved normaltilnærming bli:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

Dette er formelen for standardavvik i en binomialfordeling, men gjelder også der vi har en dikotom variabel. Når vi skal sammenligne med en annen gruppe elever med de samme forutsetninger ved hjelp av samme oppgave, spør vi om den forskjell vi finner er statistisk signifikant eller ikke. Har vi $a$ riktige besvarelser av $n$ antall besvarelser av oppgaven og sannsynligheten $p$ er satt til andelen av riktige svar, vil $z-$ verdien fortelle oss hvor mange standardavvik vi er fra sannsynligheten $p$ .

$$z = \frac{x-p}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{a}{n}-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} = \frac{\frac{a}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

der variansen til variabelen $\sigma^2 = p(1-p)$

Det er innlysende at vi får den største variansen på en oppgave når halvparten av elevene svarer riktig på oppgaven[9]. Da får vi $p = 0{,}5$ og variansen $\sigma^2 = 0{,}25$

Det er mulig å beregne konfidensintervall og å utføre eventuell hypoteseprøving. Spørsmålet er om den forskjell i gjennomsnittsverdi grupper av elever skårer på en oppgave er statistisk signifikant, eller om den kan tilskrives tilfeldigheter som skyldes utvalget. Signifikansnivået[10] oppgis alltid på forhånd, i de fleste tester settes det til 5 %, og er et mål for den usikkerhet som aksepteres ved de fleste slike undersøkelser.

Vi vet at det estimatet jeg kommer fram til sikkert ikke er identisk med den $p$ vi ville hatt dersom et større antall elever var med på undersøkelsen. Vi kan imidlertid bestemme et slingringsmonn. Det vil si at vi ut fra estimatet bestemmer et intervall som $p$ skal ligge innenfor, et såkalt konfidensintervall. Bredden på dette konfidensintervallet avhenger av hvor sikre vi vil være på at $p$ ligger i dette intervallet. Med et signifikansnivå på 5 % har vi en konfidenssannsynlighet på 95 %. Det betyr at det er 95 % sannsynlig for at konfidensintervallet dekker den riktige verdien for $p$ .

Ønsker vi å fastsette et konfidensintervall for $p$ , har vi ingen verdi for $p$ som vi kan sette inn i uttrykket for standardavviket $\sigma$ . I stedet setter vi inn estimatet og får da et tosidig konfidensintervall:

---

[9] Deriverer vi uttrykket $\sigma = \sqrt{p(1-p)}$ , får vi maksimalverdi for $\sigma$ når $p = 0{,}5$

[10] Et uttrykk for den usikkerheten vi kan akseptere ved testen. Sagt med andre ord: hvor mange prosent sannsynlighet for at nullhypotesen er sann, forutsatt at vi fortsatt tror på nullhypotesen.

$$p = \hat{p} \pm z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

der $\alpha$ angir signifikansnivå og som tidligere nevnt oftest settes til 5 %

Det er her satt som en forutsetning at de elever som er med på undersøkelsen har hatt tilnærmet den samme type undervisning med samme læreplan og samme vektlegging på berørte områder i faget. Elever med et annet undervisningsopplegg vil muligens oppnå andre verdier for $p$ og vil da få et annet konfidensintervall. Hvis det viser seg at disse konfidensintervallene ikke dekker hverandre, er dette en indikator på at gjennomsnittsprestasjonen for de to populasjonene er signifikant forskjellige. I store internasjonale undersøkelser, som for eksempel PISA, er en av hovedhensiktene å analysere utdanningspolitiske særtrekk som forårsaker slike variasjoner, og derfor er det nødvendig at utvalgsprosedyren følges for å sikre at testen ble gjennomført likt i alle deltagende land.

### 4.1.3 Analyse av to variable

Både kovarians og korrelasjon er mål for lineær samvariasjon. Dette benyttes for å måle eventuelle sammenhenger mellom to variabler. Det er ikke er lett å avgjøre om en kovarians er en indikasjon på hvorvidt det er en sterk sammenheng mellom to variable, men hvis vi dividere kovariansen (Cov) med standardavvikene kommer dette klarere frem. Vi får da en korrelasjonskoeffisient, r, også kalt Pearson product moment correlation coefficient. Korrelasjonen beregner vi med formelen

$$r = \frac{\sum_{i=1}^{n}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1 - \bar{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2 - \bar{x}_2)^2}} = \frac{Cov(x_1, x_2)}{s_1 s_2}$$

der $s_1$ og $s_2$ er standardavvikene for henholdsvis $x_1$ og $x_2$.

Korrelasjonen blir ofte kalt standardisert kovarians, og vil anta verdier mellom –1 og 1, der henholdsvis –1 og 1 indikerer perfekt negativ og perfekt positiv sammenheng. Strengt tatt bør variable være på intervall- eller forholdstallnivå, men dette følges ikke alltid. Dersom den ene variabelen er dikotom og den andre på intervall- eller forholdstallnivå, får vi "Point-biserial" korrelasjon. (se p. 4.1.4)

Sammenheng mellom variabler er ofte gjenstand for testing, og det er utviklet mange statistiske metoder for å undersøke og teste slike sammenhenger. En av de vanligste metodene er regresjonsanalyse. Denne metoden forutsetter at begge variablene er kontinuerlige, da enten på intervall- eller forholdstallsnivå, og den har derfor et begrenset anvendelsesområde i min analyse. Hvis variablene er på nominalnivå (dikotome variable er på nominalnivå), kan Kji-kvadrattest brukes, men denne testen forutsetter at antall observasjoner er større enn ti. Testen er ikke valid dersom antall observasjoner er for lite, dvs. at alle cellene bør inneholde minst fem elever. Dette følger av at Kji-kvadratfordelingen bygger direkte på normalfordelingen.

Denne testen er velegnet ved betraktning av svarmønsteret mellom to oppgaver for å se om de som svarer riktig på et spørsmål også svarer riktig på beslektede spørsmål. Et annet eksempel der denne testen kan brukes, er om det er signifikant forskjell på andelen gutter og jenter som svarer riktig på en bestemt oppgave (se problemstillingen p.1.1.1).

Kji-kvadrattesten gjennomføres etter følgende plan:

Nullhypotesen og den alternative hypotesen formuleres, og et signifikansnivå bestemmes. En tabell angir kritiske verdier for respektive signifikansnivåer.

$H_0$ : *Det er ingen forskjell*, dvs. nullhypotesen

$H_1$ : *Det er forskjell*, dvs. alternativ hypotese

Hvis nullhypotesen er riktig, vil de elevene som har besvart oppgaven riktig fordele seg forholdsmessig på begge kjønn. Siden Kji-kvadratfordelingen bygger på normalfordelingen vil kravet til en forholdsmessig fordeling bli større desto flere observasjoner vi har med.

### 4.1.4 "Point-biserial" korrelasjon

Den mest anvendte metoden som forteller oss i hvilken grad prestasjonen på en enkel oppgave (dikotom variabel der svaret er enten riktig eller galt) forholder seg til hele prøven, er "point-biserial" korrelasjonen. Man går fram på samme måte som ved utregning av *Pearson product moment correlation coefficient*, men her er bare den ene variabelen på enten intervall- eller forholdstallnivå. Den andre variabelen er som nevnt dikotom, der riktig svar premieres med 1 og uriktig gis 0. Koeffisienten angir derved i hvilken grad enkeltoppgaven måler det samme som prøven som helhet. Den gir et tall mellom –1 og 1, der tallet null representerer ingen relasjon.

En matematisk formel for utregning av "point-biserial" korrelasjon er:

$$\rho_{\text{pbis}} = \frac{m_1 - m_0}{s_x}\sqrt{p_0 p_1}$$

der

$m_0$ = gjennomsnitt for hele prøven for de som ikke svarte riktig på oppgaven

$m_1$ = gjennomsnitt for hele prøven for de som svarte riktig på oppgaven

$s_x$ = standardavvik for hele prøven

$p_0$ = andelen av de som ikke svarte riktig på oppgaven

$p_1$ = andelen av de som svarte riktig på oppgaven

Ett av formålene ved å finne denne koeffisienten, er å se om en enkeltoppgave korrelerer med prøven som helhet. Det forventes at de som svarer riktig på én oppgave, også skårer høyere på hele prøven enn de som ikke svarer riktig. Vi sier da at oppgaven har bra diskrimineringsevne. Oppgaver med lav diskrimineringsevne bør utelates ved senere tester, dersom hensikten med testen er å rangere elevene.

Ved diskusjonen av de enkelte oppgavene vil bl.a. ”point-biserial” korrelasjonen bli benyttet.

### 4.1.5 Standardisert totalskåre

Hvor mange standardavvik over eller under gjennomsnittet en elevs totale skåre ligger, uttrykkes ved z-skåre. Ett av formålene ved den gjennomsnittlige z-skåre er å se om de elevene som svarer riktig på en enkeltoppgave, også har en høyere totalskåre på hele prøven enn de som ikke svarer riktig, og ikke mindre viktig at de som ikke svarer riktig, eventuelt har en negativ gjennomsnittlig totalskåre. Viser det seg at en oppgave i pilottesten løses riktig av flere elever med lav totalskåre mens noen av elevene med høy totalskåre ikke løser oppgaven, er den lite egnet. Den bør da enten omarbeides eller droppes. Dette vil i høyeste grad gjelde dersom oppgaven skal benyttes i en undersøkelse som PISA-undersøkelsen

En matematisk formel for å regne ut den gjennomsnittlige standardiserte totalskåre for elever som svarer riktig på en enkeltoppgave er:

$$Z(1) = \frac{m_1 - m_x}{\sigma_{x-1}} \rightarrow \frac{m_1 - m_x}{s_x}$$

der

$m_1 =$ gjennomsnitt for hele prøven for de som svarte riktig på oppgaven

$m_x =$ gjennomsnitt for hele prøven for alle elevene

$\sigma_{x-1} =$ standardavvik for hele prøven[11]

$s_x =$ standardavvik for hele prøven

En tilsvarende formel kan benyttes for de elevgrupper som ikke svarer riktig på oppgaven eller ikke svarer.

Tar vi z-skåre for gruppene elever som svarer riktig, galt eller ikke svarer på en oppgave og multipliserer med de respektive antall, vil gjennomsnittet bli 0 med standardavviket 1.

Som man ser får man tilnærmet samme informasjon ved å regne ut gjennomsnittlig z-skåre som man får ved ”point-biserial” korrelasjon. Fordelen med å beregne z-skåren er at den oppgir hvor mange standardavvik enten over eller under gjennomsnittet på hele prøven eleven får enten det er svart riktig, galt eller ikke svart på en oppgave.

[11] Dataprogrammet SPSS benytter standardavviket $\sigma_{x-1}$ som er mer presist enn $s_x$ med et mindre antall elevkandidater. Denne effekten skyldes at først må vi estimere gjennomsnittet, og da har vi x-1 antall variable igjen. Med økende antall elevkandidater, vil forskjellen på standardavvikene $s_x$ og $\sigma_{x-1}$ bli ubetydelige.

## 4.2 Reliabilitet

Validitet (se p 4.3) er et spørsmål om i hvilken grad vi måler det vi har til hensikt å måle, dvs. graden av gyldighet. Reliabilitet vil på sin side, si noe om hvor godt vi måler det vi faktisk måler, dvs. graden av pålitelighet. En høy validitet forutsetter en høy reliabilitet. Reliabiliteten handler i tillegg til hvorvidt målingene er pålitelige om målingene har en intern konsistens. Høy reliabiliteten betyr at uavhengige målinger skal gi tilnærmet identisk resultat. I denne pilottesten er det et spørsmål om i hvor stor grad en elevs testskåre forblir den samme hvis vi gjennomfører en alternativ test. For å bestemme reliabiliteten kan det være ønskelig gjennomføre testen to ganger. Korrelasjonen mellom elevbesvarelsene på disse to testene ville gitt oss en reliabilitets-koeffisient. Å gjennomføre to slike tester kan ofte ha praktiske ulemper som for eksempel at det er mer arbeidskrevende. Dessuten er det fare for at elevene vil huske sine gamle svar dersom den andre testen var en gjentagelse av den første.

Denne pilottesten baserte seg på at elevene fikk én test. Det er da også flere andre metoder å regne ut reliabilitetskoeffisienten på. En mye brukt metode er *Split-half* der spørsmålene i pilottesten deles i to og man regner ut korrelasjonen mellom disse to halvdeler. Derav metodens navn.

> *Other internal-consistency measures of reliability do not require splitting the test into halves and scoring each half separately. These procedures assess the inter-item consistency, or homogeneity, of items. (Ary 1996, s.283)*

Det er ulike metoder å bestemme reliabiliteten på. Cronbachs alfa er en av disse. Det er flere faktorer som kan influere på tokningen av Cronbachs alfa, blant annet følgende:

> *Reliabiliteten er avhengig av antall oppgaver i en test. Jo flere oppgaver, desto høyere reliabilitet.*
>
> *Fordi Cronbachs alfa baseres på variansen og kovariansen til testen, spiller homogenitet i elevgruppen en rolle for en tests reliabilitet. Jo større kovarians, desto større reliabilitet.*
>
> *Testtiden spiller også rolle for reliabiliteten. Hvis en ikke blir ferdig med alle oppgavene i en test, er det åpenbart at prestasjonen på de ikke fullførte oppgavene blir fullstendig konsistent, uavhengig av om oppgavene er homogene i innhold eller ikke. (Angell 1996, s.87)*

Det vil med andre ord si at både antall spørsmål og fellestrekk ved besvarelsene vil innvirke på Cronbachs alfa. Et testresultat som er basert på spørsmål som de aller fleste elevene ikke besvarer enten fordi de oppfattes for vanskelige eller på grunn av tidspress, vil medføre at prestasjonen på de resterende spørsmål blir konsistent. Derfor er det viktig å påse at både testtiden og spørsmålenes vanskegrad er tilpasset den elevgruppen den er ment for (se også reliabilitetsbetraktninger i kapittel 8).

### *4.2.1 Reliabilitetskoeffisient*

Det er utviklet teknikker for å måle intern konsistens. Reliabiliteten er et uttrykk for den gjennomsnittelige korrelasjonen mellom alle oppgavene og antallet oppgaver, og den kanskje mest kjente måleteknikken for dette, er Cronbachs alfa. Denne metoden gir den gjennomsnittelige korrelasjonen mellom den utførte testen og alle tenkelige tester med samme antall oppgaver i et univers av tester. Man får da en indikasjon på hvilke oppgaver som er lavt korrelert med andre oppgaver, og som derfor bør utelates ved en senere test. Vi kan betrakte Cronbachs alfa som en nedre grense for reliabilitet. Matematisk er den formulert slik:

$$\alpha = \left(\frac{K}{K-1}\right)\left(\frac{s_x^2 - \sum s_i^2}{s_x^2}\right)$$

der

$K =$ antall oppgaver i testen

$\sum s_i^2 =$ summen av variansen for alle testens oppgaver

$s_x^2 =$ variansen til hele testen

En regel som ofte legges til grunn er at $\alpha$ bør være større enn 0,7, men ikke for nær 1. For høy $\alpha$ tyder på at spørsmålene i testen er for like eller at det er unødvendig mange, og er derfor en indikator på at det er grunn til å stille spørsmål ved om testen virkelig måler det vi ønsker. Skulle $\alpha$ bli for lav kan vi øke testens reliabilitet ved å øke antall oppgaver, slik vi kan se ut fra formelen hvor antall oppgaver inngår. Med en Cronbachs alfa på 0,913 for denne pilottesten (se reliabilitetsbetraktninger kap.8), vil det bety at 91,3 % av variansen er felles for sann og observert skåre, og 8,7 % skyldes da tilfeldigheter. (Mer om dette, se Angell og Lie 1993, s.13).

## 4.3 Validitet

En undersøkelses validitet er i hovedsak et uttrykk for i hvor stor grad undersøkelsen måler det man har til hensikt å måle. Man kan derved si at begrepene validitet og reliabilitet beskriver hvor bra undersøkelsens test har fungert.

> *Mens **reliabilitet** uttrykker hvor nøyaktig eller pålitelig vi måler det vi faktisk måler, kan vi si at **validitet** er et mål på i hvilken grad testen måler det den gir seg uttrykk for å måle. (Angell 1996, s.89)*

Selv om de slutninger og konklusjoner vi kan trekke på grunnlag av testen har høy validitet, ville den ha liten verdi hvis den ikke er generaliserbar, dvs. at analysens funn kan gjøres allmenngyldige. Den må også kunne gjelde for andre situasjoner enn den testen er gjennomført i. I motsetning til en kvalitativ undersøkelse der man kontinuerlig vurderer validiteten og reliabiliteten gjennom alle stadier i undersøkelsen, vil en i en kvantitativ undersøkelse velge en testmetode med kjent og akseptabel validitet til det formål man har. Det kan også være formålstjenlig å benytte begge metodene i én og samme undersøkelse for å få en samtidig vurdering av i hvor sterk grad testen evner å ta opp spørsmålene i en problemstilling.

Minoritetsspråklige elever og elever med lesevansker har lettere for å hoppe over oppgaver med mye tekst selv om de behersker matematikken i oppgaven. Hvis testens hensikt var å kartlegge matematikkunnskapen til eleven, kan en test med mye tekst derfor ha lavere validitet. Dersom testens hensikt var å måle elevers tekstforståelse av en matematikkoppgave for deretter å bruke sine matematikkunnskaper, ville testen ha høyere validitet. En undersøkelse der mange av oppgavene har mye tekst, vil være tjent med bruk av kvalitativ metode.

### *4.3.1 Innholdsvaliditet*

I hvilken utstrekning vil en matematikktest dekke alle sider ved et emne? En hjelp til vurdering av dette er å sette opp en liste over emnets ulike sider. En test som fanger opp det vesentlige innen emnet vil ha en høy innholdsvaliditet. En test som gjennomføres i flere ulike land vil trolig bare dekke deler av læreplanen for norske elever. Den vil derfor ha lavere innholdsvaliditet i forhold til å dekke matematikkunnskapene i henhold til L97. En eksamen bør ha høy innholdsvaliditet siden eksamensoppgaver bør dekke det vesentlige av læreplanen. Det er derfor viktig at forslag til eksamensoppgaver blir utviklet eller vurdert av erfarne lærere som til daglig arbeider med læreplanen og har god erfaring fra undervisning. Det er grenser for hvor stor del av læreplanens innhold som kan berøres i en eksamen med begrenset tid, det vil alltid bli et utvalg. En eksamensform med betydelig andel flervalgsoppgaver, vil ha større muligheter til å dekke mer av læreplanen enn tradisjonelle åpne oppgaver. Kanskje er dette et argument for å utarbeide eksamensoppgaver med et stort antall flervalgsoppgaver?

Innholdsvaliditeten til min pilottest vil primært være av interesse i forhold til rammeverket for PISA-undersøkelsen. Det vil i denne situasjonen bli en vurdering av hvor representativt pilottesten dekker dette rammeverket. Graden av representativitet utgjør testens innholdsvaliditet. Så lenge mine oppgaveforslag dekker et begrenset område av PISAs rammeverk, vil dette ha betydning for vurdering av innholdsvaliditeten. Selv om oppgavene ligger innenfor felles mål i L97, er det viktige sider av fagplanen som ikke blir dekket. Derfor vurderer jeg også innholdsvaliditeten som lav i forhold til avgangsprøven.

### *4.3.2 Kriterievaliditet*

Det sentrale i en test er at den måler det den har til hensikt å måle. Det vil si at det resultatet vi får ved en test vil overensstemme med de resultater andre tester vil få, enten testen blir utført av andre eller om det blir benyttet andre testmetoder. Derfor må den variabelen som blir målt, samvariere med en teoretisk variabel som det forventes at den vil samvariere med. Den teoretiske variabelen er da kriteriet. Det kan ofte være vanskelig og tidkrevende å definere en kriterievariabel på tilfredsstillende måte. For å fastsette kriterievaliditeten til konklusjonene av en test, er det derfor viktig på forhånd å legge mye arbeid i dette definisjonsproblemet.

Det er to typer av kriterievaliditet. Hvis kriterievariabelen foreligger samtidig som pilottesten, vil det være snakk om samtidig validitet. Det må da bli en relasjon mellom en test og et kriteriemål. Et eksempel på samtidig validitet, er å se på testresultatet og den standpunktkarakter eleven har når testen ble foretatt.

Ofte foreligger kriterievariabelen mer eller mindre lang tid etter en test. I den situasjonen vil det være snakk om prediktiv validitet. Hvis eleven viser stabilt høyt nivå i faget på grunnlag av en rekke prøver gjennom året, forventes det at en eksamen også gir tilsvarende god karakter. I hvilken grad vil en test fortelle oss hvordan eleven vil prestere på en avgangsprøve, der avgangsprøven er kriterievariabelen? Det er først etter avgangsprøven vi får vite hvordan eleven har klart seg, og derved også kan vurdere den tidligere prøvens prediktive validitet.

### *4.3.3 Construct validitet*

Ved alle typer av validitet er det forbindelsen mellom en test og en annen variabel som er av interesse. Denne variabelen vil ha forskjellige betydninger avhengig av intensjonen ved en test. I de to foregående typer validitet er variabelen forholdsvis avgrenset og konkret. Ved construct validitet kan variabelen være mer diffus, for eksempel en abstrakt psykologisk egenskap eller de begrepsoppfatninger som ofte gir seg uttrykk i kunnskaper i matematikk.

> *Et construct er et hypotetisk begrep, og eksistensen av et construct kan aldri bli absolutt bekreftet. Det er heller ikke målbart. (Kleve 1994, s.30)*

Det vil være av interesse å betrakte spørsmålene slik de er formulert i testen i forhold til dette construct. Ved en test som har som oppgave å måle elevenes dyktighet i et fag, vil det være naturlig å granske om spørsmålene fanger opp alle aspekt ved det underliggende construct. Høy validitet er en indikasjon på det.

Det er mulig å vurdere høy validitet ved både kvantitative og kvalitative betraktninger selv om det siste ofte i praksis kan by på tolkningsproblemer. En vanlig anvendt framgangsmåte er å foreta korrelasjonsundersøkelse. Dersom en korrelasjonsundersøkelse mellom testen og en annen test som erfaringsmessig måler de samme egenskaper viser høy korrelasjon, vil det også være håp om at resultatet og de konklusjoner som kan trekkes på bakgrunn av testen, har høy validitet.

I denne hovedoppgaven vil jeg benytte en kvalitativ metode for å vurdere i hvilken grad spørsmålene i pilottesten overensstemmer med elevenes begrepsoppfatninger slik de fremkommer i målområdene for L97 (se p. 2.1) og deres kunnskaper og ferdigheter slik det er betegnet i PISA-rammeverket (se p.1.3). Det er da naturlig å betrakte spørsmålene slik de er formulert i testen i forhold til disse constructer. Her vil også en analyse av testens resultater være til stor hjelp i denne vurderingen.

# 5. Oppgaveutvikling

## 5.1 Signaler fra PISA

Retningslinjer for utvikling av oppgaver til PISA-undersøkelsen er for matematikkens del gitt i **Mathematics item development for PISA 2003 and item submission guidelines** (OECD/PISA 2000,2001a & 2001b).

Rammeverket til forrige runde, PISA 2000, konsentrerte seg om to sentrale ideer, Forandring og sammenheng samt Rom og form. Neste fase, PISA 2003, inkluderer også Kvantitativ resonnement og Usikkerhet. Derved dekkes alle de fire sentrale ideene. Det var derfor et stort behov for forslag på varierte oppgaver fra de deltagende landene.

> *For PISA 2003 Mathematics, about 130 items will be needed for the Main Study, some of which can be selected from PISA 2000 item pool. The target will therefore be about double this (260 items) for the Field Trial. To achieve this about 390 items will be needed for the pilot stage, from which the Field Trial items will be selected. (OECD/PISA 2000 s.6)*

I de samme Guidelines utgitt i 2000 (OECD/PISA 2000) og 2001 (OECD/PISA 2001a og OECD/PISA 2001b) gis det også retningslinjer som settes i forbindelse med oppgaveutviklingen. Det gjelder først og fremst at hver oppgave bør innledes med en autentisk kontekst som skal presentere relevante informasjoner elevene skal benytte for å løse problemet.

> *In addition, item contexts that can be regarded as authentic will be preferred. That is, PISA values most highly tasks that could be encountered in one of a variety of real-world situations, and that have a context for which the use of mathematics to solve the problem would be authentic. (OECD/PISA 2001b s.25)*

”Helt” autentisk er det ikke alltid ønskelig eller mulig at oppgavene skal være. Eksempel på det er PISAs bruk av myntenheten *zed* fra et oppdiktet *Zedland* for at enkelte land ikke skal favoriseres ved at vedkommende lands myntenhet benyttes. Jeg oppfatter at oppgavene kan baseres på en tekst direkte hentet fra media eller annen informasjonsflom, men dette må gjøres med stor forsiktighet slik at ikke elever fra oppgavens hjemland får fordeler framfor de øvrige. Kravet innebærer at teksten skal inneholde problemstillinger fra dagliglivet. Målet er å undersøke elevenes funksjonelle kunnskaper i matematikk og deres evner til å nyttiggjøre seg disse i dagens informasjonsflom. Derfor blir det også bedt om at det tas hensyn til elevenes leseferdighet ved utvikling av oppgaver. Et dilemma er at hvis konteksten til en oppgave omskrives eller forenkles i håp om at ”flere” vil mestre oppgaven, vil det også være en fare for at det autentiske vil gå tapt. Ikke minst kan dette skje der konteksten er hentet fra et avisoppslag eller et annet media.

> *The level of reading required to successfully engage with an item should be considered very carefully. The wording of items should be as simple and direct as possible. This is an assessment of Mathematical literacy, not of reading ability. (OECD/PISA 2000 s.11)*

Videre står det:

> *In addition, high reading demands in the formulation of items can have a differential impact on boys compared to girls at least in some countries. (ibid, s.12)*

PISA 2000 viste klart at det er store kjønnsforskjeller mellom jenters og gutters gjennomsnittsskåre i lesing generelt og at dette i alle land var i jentenes favør. Noe mindre var forskjellen i retrieve-skalaen som viser i hvilken grad elevene er i stand til å hente ut de relevante informasjonene fra en tekst. Ett av målene i PISA er å undersøke elevenes evner til å nyttiggjøre seg matematikken i dagliglivet. Jeg anser at det derfor også bør være aktuelt å undersøke forskjellen på jenters og gutters prestasjon på matematikkoppgaver nettopp basert på autentiske kontekster (se problemstillingen p.1.1.1).

Ved oppgaver som skal benyttes i et stort antall land med forskjellige normer og holdninger, er det viktig at konteksten i oppgavene verken virker støtene på grupper av elever eller favoriserer enkelte land ved at elever i noen land føler seg mer fortrollig med problemstillingen.

> *Care should be taken to avoid question contexts that would create a cultural bias. Keep in mind that the test will be administered in a large number of countries with big cultural and geographical differences. (ibid)*

I mitt første utkast til oppgaver som ble presentert for ILS var det med en oppgave rundt Body- Mass- Index (BMI) som behandlet forholdet mellom høyde og vekt ved utregning av indeksen. I ILS mente man det var fare for at noen av elevene ville reagere emosjonelt på problemstillingen, oppgaven ble derfor droppet. Her tok man både hensyn til den aldersgruppen oppgavene ble utviklet for og eventuelle holdninger i enkelte av de deltagende land, noe jeg har forståelse for. En oppgave med samme tema ble for øvrig gitt til avgangsprøven i matematikk i grunnskolen for de voksne.

I en internasjonal undersøkelse som den PISA gjennomfører er det viktig at elever fra enkelte land ikke skal ha noen fordeler som at de enten kjenner oppgaven fra tidligere, eller er mer fortrolig med problemstillingen.

> *To avoid unfairly advantaging students from any particular country, items should not be directly extracted from textbooks or other common resource materials routinely used by students in any country. (ibid, s.11)*

Dette er et selvfølgelig, men vanskelig krav. En oppgaveprodusent vil som oftest kjenne til de oppgaver som elevene får prøve seg på i sitt eget land, enten det er tidligere gitte eksamensoppgaver eller de er hentet fra lærebøker eller oppgavesamlinger. Det vil være praktisk talt umulig å ha tilsvarende oversikt over oppgaver som benyttes i alle de andre landene. Det vil derfor kunne utvikles oppgaver som elever i andre land lettere vil mestre. Et eksempel på det er oppgaven Fahrenheit-Celsius. (se p.7.4) I land som ofte benytter fahrenheitskalaen selv om celsiusskalaen er den offisielt vedtatte, vil elevene ha klare fordeler. Oppgaven ble derfor ikke tatt med i PISAs Generalprøve. I Norge vil jeg anse at denne oppgaven, muligens i noe omarbeidet form, vil fungere bra som en testoppgave.

Et meget viktig krav til oppgaver er:

> *In some cases, PISA items will be developed in "Units" that comprise a number of questions (perhaps 3 to 5) relating a single stimulus. Such questions should as far as possible be independent of each other in the sense that answer to one question should not depend on the student having answered a previous question. (ibid)*

Alle oppgaveforslagene til pilottesten (presentert under kapittel 7) inneholder fra fire til seks spørsmål. Flere av spørsmålene i en oppgave er forholdsvis like enten ved at de både er presentert som åpen oppgave eller flervalgsoppgave. En av hensiktene med pilottesten var blant annet å se hvilke spørsmål som ville fungere best som testoppgave for siden eventuelt å droppe de andre. Resultatet av pilottesten ga også opplysninger som kunne indikere om det ville være aktuelt å slå sammen flere spørsmål til en ny og forhåpentligvis bedre testoppgave.

Det kan by på større utfordringer å utvikle oppgaver med flere spørsmål. Med hensyn til testens pålitelighet er det viktig at spørsmålene er formulert slik at forutsetningen for å svare riktig på et spørsmål ikke er avhengig av om eleven har mestret noen av de andre spørsmålene i oppgaven. Dette krav har jeg ikke helt fulgt i alle oppgavene til pilottesten. Årsaken var at pilottestens intensjon var å se hvilke spørsmål i en oppgave som kunne fungere best til PISA-undersøkelsen.

Det bør også etterstrebes at spørsmål med forventet høy p-verdi presenteres først i en oppgave. Derved minsker man faren for at noen elever tidlig blir "skremt" fra å prøve å besvare de andre spørsmålene i oppgaven. Dette vil være vanskelig å praktisere i PISA siden en oppgave som kommer tidlig i ett testhefte, kan dukke opp på slutten i et annet hefte.

## 5.2 Endelige oppgaver til pilottesten

Ved utvikling av de endelige oppgavene til pilottesten hadde jeg satt meg som mål både at alle de fire Sentrale ideer og de fem Situasjoner – kontekst ble representert i testen. Dette oppnådde jeg i hovedsak med unntak av oppgaver fra Skoletilværelsen. Spørsmålene er hentet fra alle Kompetanseklasser med hoveddelen fra klasse 2. Å fastsette kompetanseklasser for noen av spørsmålene etter de kriterier som er gitt kan være noe problematisk siden det ikke er klare grenser mellom de tre kompetanseklassene.

Jeg har ved oppgaveutviklingen så langt som mulig fulgt de retningslinjer som er presentert og kommentert under punkt 5.1, kanskje med unntak av forutsetningen om å kunne svare på et spørsmål uten at eleven har mestret de foregående i oppgaven. Eksempel på det er oppgaven Fibonaccitallene. Hvis ikke elevene mestrer mønsteret av kvadratene på første spørsmål, vil de ikke ha noen muligheter for å svare riktig på de andre spørsmålene. Jeg minner her om at hensikten var å teste oppgavene, og pilottesten ville legge grunnlag for hvordan den endelige oppgaven ville fremkomme enten ved å fjerne noen spørsmål eller forandre på noen.

Jeg har også lagt vekt på at oppgavene skulle ligge nær opp til det elevene møter i skolen. I en omarbeidet form vil de kunne fungere som eksamensoppgaver til avgangsprøven. Selv i den siste oppgaven, Totallsystemet (se vedlegg), hvis emne ikke ligger innenfor læreplanverket for den 10-årige grunnskolen, ble hensynet ivaretatt ved at innledningen ga en grundig, men nok noe for omfattende forklaring på hvordan man konstruerer totallsystemet. En av hensiktene med en slik oppgave var å undersøke elevenes evne til å besvare spørsmål på et tema de trolig kjente lite til før oppgaven ble presentert. Den ble luket ut av PISA tidlig i prosessen. Årsaken kan være at elever i noen deltagende land kjenner godt til omregning fra titallsystem til totalsystem og disse ville uttvilsomt ha en fordel under større internasjonale tester som den PISA gjennomfører. Dette var en oppgave jeg utviklet etter ønske fra ILS som en erstatning for oppgaven BMI som ble droppet tidlig (se p.5.1).

Alle oppgavene, unntatt *Totallsystemet*, er presentert og analysert i kapittel 7.

## 5.3 Kodesystemet

Siden de respektive svaralternativene på spørsmålene skal registreres og analyseres i et statistikkprogram, må hvert av dem få en kode. For de åpne oppgavene var det ønskelig å benytte samme kodesystem som tidligere var benyttet i PISA-undersøkelsen våren 2000 og som hadde høy pålitelighet. Besvarelsene for de åpne oppgavene ble derfor i første omgang gitt ensifrede koder. For **Riktig svar** ga jeg kode 1 eller 2 avhengig av det arbeid som lå bak besvarelsen. Ble kode 2 benyttet, ville kode 1 representere delvis riktig svar. Kodene for **Ikke riktig svar** og **Ikke svart** var henholdsvis 0 og 9. For disse oppgavene ble denne kodingen gjort umiddelbart etter testen for at resultatene av pilottesten skulle foreligge så tidlig som mulig.

Jeg innså etter hvert at kodesystemet var mangelfullt. Mye informasjon gikk tapt, ikke minst for **Ikke riktig svar**. Jeg kodet derfor svarene på nytt med tosifrede koder for å ivareta den diagnostiske informasjonen som feil svar gir (en av mineproblemstillinger). Det var da naturlig å benytte et godt gjennomprøvd kodesystem som blant annet ble benyttet for Third Internatonal Mathematics and Science Study (TIMSS).

> *For å ivareta den diagnostiske informasjonen som elevens svar gir, ble det utviklet et tosifret kodesystem for hver oppgave. Det første sifferet sier om svaret er riktig eller galt eller eventuelt delvis riktig. Det andre sifferet refererer til typer av svar, for eksempel hvilke tanker de har om et begrep eller hvilken metode de har brukt i løsningen. Koder mellom 20 og 29 vil si at det er gitt to poeng for svaret, men koder mellom 10 og 19 vil si ett poeng. Koder mellom 70 og 79 viser at svaret er feil. Det ble gitt en egen kode på de oppgavene der elevene ikke hadde svart.*
> *(Brekke, Kobberstad, Lie & Turmo 1999, s.15)*

Spørsmål som ikke ble besvart, er i denne pilotundersøkelsen kodet 99.

Mange av de åpne oppgavene til pilotundersøkelsen ble presentert uten krav om forklaring eller om å vise utregning. Dersom det da oppgis feil svar til et spørsmål, vil det være vanskelig å få innsikt i hvordan eleven hadde tenkt eller om svaret beror på en regnefeil. Med tosifrede koder var det mulig ved hjelp av krysstabeller å følge elevens svarmønster på flere spørsmål. Dette gir større innsikt i hvorvidt eleven har brukt feil strategi eller om det er misoppfatninger som ligger til grunn.

I tillegg til koding av åpne oppgaver, måtte flervalgsoppgavene også få sine koder. Til de enkle flervalgsoppgavene ble svaralternativ A kodet 1, B kodet 2 osv. Det riktige alternativet ble i tabellen merket med *.

For de sammensatte flervalgsoppgavene, spørsmålene 8, 30 og 31 ble **Ja** kodet 1 og **Nei** kodet 2. Det riktige alternativet ble også her merket med * i tabellene.

# 6. Utprøving

## 6.1 Hvilke skoler

For å sikre en relativt høy reliabilitet på testen på tross av eventuelt frafall, var det ønskelig at antall klasser/grupper som var med på uttestingen av oppgaveforslagene var slik at antall elever ble rundt 150. Det var viktig at antall elever var stort nok til at også oppgaver med høy vanskegrad ble besvart i uttestingen. Med et så stort antall elever ville også estimeringen av oppgavenes p-verdi (vanskegrad, se p.4.1.1) ligge nærmere den p-verdien vi ville fått dersom hele populasjonen, det vil si alle 10-klassinger, var med på testen. Dessuten burde utvelgelsen av skoler/klasser skje med et sannsynlighetsutvalg på grunnlag av skolenes størrelser med eventuelt flere parallelle klasser eller skolens regionale beliggenhet. Siden pilottesten ikke hadde som oppgave å måle elevenes kunnskaper i matematikk, men hvordan den enkelte oppgave fungerte i testen, ble den av praktiske grunner gjennomført uten strengt å følge de regler som gjelder for et sannsynlighetsutvalg.

For å ha en viss spredning av skolene ble det derfor av praktiske grunner tatt kontakt med tre skoler i og rundt Oslo, en sentrumskole i Oslo, en skole i Oslo øst og en skole vest for Oslo. Disse ble valgt blant annet fordi elevsammensetningen og antall 10-klassinger var forskjellig for disse skolene. Disse skolene var også kjent med tidligere PISA-undersøkelser. To av skolene deltok i hovedtesten i 2000 og én deltok i generalprøven i 1999. På det tidspunkt testen ble gjennomført var klassene/gruppene stort sett i gang med repetisjon og forberedelse til avgangsprøven. Testen ble kanskje av noen elever oppfattet som en slags generalprøve før en eventuell avgangsprøve.

## 6.2 Elevsammensetning

Elevsammensetningen var noe forskjellig på de tre skolene. På skolen sentralt i Oslo var det det skoleåret testen ble gjennomført, bare én 10-klasse, mens den normalt har hatt to parallelle klasser. I denne klassen var det 9 jenter og 10 gutter som gjennomførte testen. Av disse hadde to minoritetsspråklig bakgrunn.

Skolen vest for Oslo er en relativt stor ungdomsskole der elevene rekrutteres fra et stort geografisk område. Fire av fem 10-klasser/grupper på skolen det året deltok i uttestingen, og antallet jenter var relativt høyt. Den klassen/gruppen som ikke deltok, hadde tilsvarende høy andel gutter. Av de 82 elevene som deltok ble det registrert 50 jenter og 28 gutter ved deres egen anførsel på besvarelsen. Ved senere undersøkelse ble det konstatert at de fire som ikke påførte kjønn på oppgaven måtte være gutter så det totale antall gutter ble 32. Under 3 % av elevene hadde minoritetsspråklig bakgrunn.

På skolen i Oslo øst var det 20 jenter og 22 gutter som deltok på testen. En av tre klasser reserverte seg grunnet tidsnød i forbindelse med manglende tid til repetisjon før en eventuell avgangseksamen. En vesentlig del av elevene ble rekruttert fra et gammelt etablert villaområde slik at andelen minoritetsspråklige elever var noe mindre enn ventet, noe over 10 %. Dette er en andel betraktelig lavere enn andre nærliggende skoler som kan ha opptil 50 % med minoritetsspråklig bakgrunn.

## 6.3 Praktisk gjennomføring

Av timeplantekniske grunner var det ikke mulig å gjennomføre alle testene tidlig på dagen, noe som var ønskelig. På to av skolene ble testen gjennomført i de to første timene på dagen, mens på skolen i Oslo øst gjennomførte den ene gruppa testen noe senere på dagen. Her lå det en fare for at elevenes form og motivasjon ble noe redusert. I tillegg var det en liten, men tilstedeværende, fare for at elevene ville samtale om testen. Slike tilfeldigheter ville eventuelt påvirke testresultatet. Denne faren som lå i at de elevene som gjennomførte testen tidligere på dagen kunne gi hint til den siste klassen som skulle ta testen senere var liten siden tidsavstanden mellom begge gjennomføringene var minimal. Med større tidsavstand ville den vært tilstede. I den siste klassen ble det observert at noen elevers motivasjon ble redusert på slutten av testen. Flere ønsket å avslutte før testen var ferdig. Dette bekrefter at slike tester bør gjennomføres så tidlig som mulig på dagen og samtidig for de klasser/grupper som deltar.

Presentasjon av testen ble i hovedsak godt mottatt av elevene selv om det viste seg at både motivasjon og innsats varierte en del. Elevene fikk i forkant en kort orientering om testens hensikt og at det ikke er klassens resultater som skal vurderes, men at det er oppgavene som skal bedømmes for eventuelt å kunne brukes i en større internasjonal undersøkelse. Dessuten ble de bedt om å lese innledningen til hver oppgave grundig før de svarte på de respektive spørsmål, siden viktige informasjoner for å løse oppgavene ligger i teksten. Elevene ble også informert om at de kunne gi et signal dersom det var noen uklarheter i oppgavesettet. Flere av elevene var ikke kjent med flervalgsoppgaver og det ble derfor presisert at de skulle velge det eller de svarene som passet best. Her var det viktig hvis mulig å unngå gjetting selv om det også var ”tillatt”.

Det var tydelig at ikke alle elevene satte like stor pris på flere av de typene oppgaver de møtte i uttestingen. Det viste seg at møtet med den innledende teksten på noen av oppgavene skapte usikkerhet hos noen av elevene. Dette gjaldt spesielt der konteksten i en oppgave var situasjoner lokalisert lengst bort fra elevens erfaringsverden (se p. 1.3.3). Eksempler på det er *Fibonaccitallene* og *Totallsystemet* (se vedlegg). Den sistnevnte var den siste oppgaven i oppgavesettet. Den første oppgaven, *Potetoppgaven*, hadde en problemstilling fra dagliglivet som i avstand lå nær eleven (*Personlig liv*). Den innledende teksten virket tilsynelatende problemfri for elevene, og dette ble oppfattet av noen som en ”lett” oppgave. Det kan være mange grunner til at en test eller prøve bør starte med oppgaver de fleste ikke har de store problemene med å løse. Mer om dette i p.7.1.1.

Oppgavesettet ble avsluttet med et ”*Innsatstermometer*” der elevene ble bedt om krysse av for hvor stor innsats de selv mente de hadde lagt ned i den testen som de nettopp hadde jobbet med. Inntrykket etter både samtaler med lærere og elever, var at flere oppfattet denne skalaen som et uttrykk for hvordan de antok at de hadde gjort det på testen.

”*Innsatstermometeret*” fungerte derved neppe optimalt etter intensjonen. Av to grunner burde det da også vært med en skala på hvordan elevene følte de selv mente de hadde prestert på testen. For det første ville elevene klarere se forskjell på innsats og prestasjon. I tillegg ville det være av stor interesse å undersøke hvordan elevene følte de hadde prestert på denne testen sett i forhold til faktisk resultat.

”*Innsatstermometeret*” har en skala fra 1 til 10, der 10 er høyest innsats. Det vil være av interesse å se på samvariasjonen mellom poengskåre og elevenes ”innsats”. Til dette kan Pearson product moment correlation coefficient presentert i p. 4.1.3 benyttes. Denne korrelasjonskoeffisient gir et mål på elevenes oppnådde poeng i forhold til ”Innsatstermometeret” og gir informasjon om samsvar mellom prestasjon og innsats (se konklusjonen kapittel 9). Det vil foreligge en usikkerhet ved en slik sammenlikning avhengig av påliteligheten av elevenes avkryssing, også om vi hadde vært sikre på at alle elevene hadde forstått ”*Innsatstermometeret*” korrekt. Elever som føler de her presterte dårlig på en test, vil kanskje krysse av på at den innsats de gjorde på testen, var liten.

# 7. Analysen

I utgangspunktet valgte jeg å drøfte de 6 oppgavene som PISA selv ønsket å utteste. Etter den statistiske analysen av alle 8 oppgavene antok jeg at svarmønsteret på oppgaven Kast med to terninger også burde drøftes nærmere. Analysen viste at denne oppgaven ikke svarte til de ønsker PISA satte for testoppgaver, men besvarelsene avdekket andre interessante sider ved oppgaven.

De 7 oppgavene med til sammen 33 spørsmål som jeg drøfter nærmere, blir presentert i dette kapitlet. Den siste oppgaven, *Totallsystemet*, blir presentert i vedlegg med både frekvens, prosent og utregnet z-skåre og "point-biserial" på hvert av spørsmålene. Elevbesvarelsene er bearbeidet etter de retningslinjer som er nevnt under kapitlet om testteori (se p.4.1.1 – 4.1.5). Hvert av spørsmålene blir etterfulgt av en tabell som angir svaralternativene. I tabellen har jeg også tatt med frekvens og prosent for svaralternativene. I tillegg er det utregnet både gjennomsnittlig z-skåre og "point-biserial". En z-skåre som er regnet ut på grunnlag av få besvarelser, sier lite. Derfor blir det oppgitt "point-biserial" korrelasjon for hvert svaralternativ. C. Angell sier i sin Dr. scient-avhandling:

> *Høy" point-biserial" gir som regel også høy gjennomsnittlig totalskåre. Men hvis det er få elever i en gruppe, kan gjennomsnittskåren godt være høy selv om det ikke er høy samvariasjon, altså høy "point-biserial". "Point-biserial" er regnet ut på en måte der det tas hensyn til antallet og spredningen i utvalget, noe en ikke gjør når en bare regner ut en gjennomsnittverdi.(Angell 1996, s.146)*

For flere av spørsmålene har jeg benyttet krysstabeller for å undersøke nærmere svarmønstre på beslektede spørsmål for dermed å sammenfatte viktige funn som blir kommentert i oppsummeringen min etter hver av oppgavene.

## 7.1 Oppgaven Poteter

Ved utvikling av matematikkoppgaver til PISA var uttesting av oppgavene sentralt for å se om de egnet seg til en større internasjonal test. I denne uttestingen var det viktig å ha med oppgaver elevene var motivert for å løse. Mange års erfaring som lærer både i klassesituasjonen og ved betraktning av elevprestasjoner på prøver, tester og eksamener, har gitt meg en erkjennelse av at flere elever blir lettere tent på å løse oppgaver som de umiddelbart føler de har tak på selv om de testteoretisk ikke alltid er ideelle[12]. Det er også viktig at prøvene starter med oppgaver som forventes å ha høy p-verdi, dvs. lav vanskegrad. Det beste er om slike oppgaver også kan dukke opp senere i prøven. Dette fordi elevene ikke skal oppfatte at testen blir vanskeligere etter hvert som de løser oppgavene. Hvis testen har en progresjon med oppgaver der vanskegraden etter hvert blir høyere, vil flere elever føle at de har nådd en grense og dermed gi opp.

[12] Oppgaver med for høy eller for lav vanskegrad vil i en testsammenheng være kostbare oppgaver. Det "ideelle" vil være p-verdi rundt 0,5. Dersom målet er å diskriminere mellom de flinkeste og de andre, vil en lav p-verdi være ønskelig. Hvis målet er å skille ut de svakeste, vil derimot en høy p-verdi være ønskelig.

En intensjon med den første oppgaven, *Poteter*, var nettopp å gi elevene en god start. Oppgaven hadde en enkel innledende tekst og enkle problemstillinger de aller fleste kjenner. Konteksten var fra en situasjon som i distanse står elevene nærmest, *Personlig liv*. For å besvare de tre første spørsmålene måtte elevene bruke *Kvantitativ resonnement*, som også er en viktig forutsetning for å foreta enkle beregninger i dagliglivet dersom man skal fungere i samfunnet. Det er også etter min oppfatning naturlig å plassere spørsmålene i kompetanseklasse 1.

De to siste spørsmålene i oppgaven faller under det som i rammeverket for PISA blir definert som *Forandring og sammenheng*. Disse to spørsmålene ble gitt som flervalgsoppgaver der elevene skulle bestemme hvilke funksjonsuttrykk og graf som dekket de opplysninger den innledende teksten ga. Dette er to spørsmål som henger mye sammen. De blir derfor også analysert ved hjelp av krysstabell. De er vurdert til å ligge i kompetanseklasse 2.

Det må også legges til at de to siste spørsmålene strengt matematisk har sine svakheter. Funksjonene vil aldri være kontinuerlige som følge av det fungerende myntsystem med avrundingsregler. I håp om å unngå dette problemet fikk spørsmål 4 følgende formulering: *Hvilket av disse utrykkene viser denne sammenhengen*? Forhåpentlig ville spørsmålsformuleringen nøytralisere problemet. Det ser i hvert fall ikke ut som det skapte noen problemer under testen.

Ideen til denne oppgaven har jeg fra mange års praksis med introduksjon av lineære funksjoner ved hjelp av en praktisk oppgave. Hvorvidt oppgaven fungerer like godt som en testoppgave kan selvfølgelig diskuteres, noe analysen av resultatet vil vise.

---

## POTETER

Mona skal handle poteter som koster kr 2,50 per *kg*. Hun har med seg 20 kroner.

Hvor mye har hun igjen når hun forlater butikken?

---

**Spørsmål 1:** POTETER

Hvilken opplysning mangler du for å kunne svare på spørsmålet ovenfor?

……………………………………………………………………………………………

---

Mange vil nok hevde at dette ikke er en ren matematikkoppgave, men min hensikt med denne starten var å gjøre elevene oppmerksom på betydningen av å lese den innledende teksten nøye.

## Koder til spørsmål 1

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | Å vite mengden poteter som ble kjøpt | 137 | 95,8 | 0,046 | 0,22 |
| | **Ikke riktig svar** | | | | |
| 70 | Teksten manglet tilstrekkelig informasjon | 1 | 0,7 | -0,193 | -0,02 |
| 79 | Forslag på beløp hun satt igjen med | 4 | 2,8 | -1,317 | -0,24 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 1 | 0,7 | -0,784 | -0,07 |

Selv tatt i betraktning at spørsmålet ligger under kompetanseklasse 1 har denne oppgaven en meget høy p-verdi på hele 0,96. Mer overraskende var det at det bare var én elev som ikke svarte på spørsmålet, noe som tyder på at intensjonen, en god start som tenner elever, ble innfridd. Ville nå oppgaven øke elevenes motivasjon til å fortsette videre på testen?

Kode 70 ble benyttet for den ene eleven som svarte at den innledende teksten manglet tilstrekklige informasjoner for å kunne besvare spørsmålet. Denne eleven hadde en negativ z-skåre like under null på hele testen. De fire besvarelsene med forslag på et beløp Mona satt igjen med når hun forlater butikken, ble kodet 79. Disse elevene var også meget svake med en gjennomsnittlig z-skåre på -1,317 og satset i hovedsak lite på å gjennomføre hele testen.

Den høye p-verdien influerer også på ”point-biserial” som blir på bare 0,22 for kode 10 (se formelen under p.4.1.4). Tatt i betraktning den svært høye p-verdien og verdien for ”point-biserial”, diskriminerte spørsmålet ikke så dårlig mellom de aller svakeste og de andre.

---

**Spørsmål 2:** POTETER

Hvor mange *kg* poteter kan hun kjøpe?

… kg

---

For å kunne svare riktig på dette spørsmålet må elevene kunne:

- vurdere spørsmål omkring personlig økonomi ved pris og kjøp
- bruke de fire regningsartene på hele tall, desimaltall og eventuelt brøk

## Koder til spørsmål 2

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 8 kg | 138 | 96,5 | 0,041 | 0,22 |
| | **Ikke riktig svar** | | | | |
| 79 | Regnefeil | 3 | 2,1 | -0,784 | -0,15 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 2 | 1,4 | -1,671 | -0,20 |

Dette er en testteoretisk kostbar oppgave med forventet høy p-verdi. Den ble da også testens høyeste, på nesten 0,97, og en svarprosent på 98,6. Til sammenligning var den gjennomsnittlige p-verdien for hele testen 0,41.

Spørsmålet er også av en type de aller fleste kjenner fra skolematematikken allerede så tidlig som på mellomtrinnet i grunnskolen. Av den grunn var det naturlig å plassere den i kompetanseklasse 1. Begrunnelsen for å ha med et slikt spørsmål er lik den som ble gitt i foregående oppgave. Nettopp fordi at så mange svarte riktig på spørsmålet ble den gjennomsnittlige z-skåren like over null. Her ser vi igjen at spørsmålet i likhet med spørsmål 1 diskriminerte bra mellom de svært dårlige og de andre med en "point-biserial" korrelasjon på 0,21 for kode 10.

De feil som ble kodet 79 var i hovedsak regnefeil, mens de to som ikke svarte, kodet 99, var testens svakeste kandidater slik den gjennomsnittlige z-skåren også viser.

---

**Spørsmål 3:** POTETER

Hvor mye har hun igjen dersom hun kjøper 4,5 *kg* poteter?

………kr

---

Her er det brukt to koder for riktig svar, 10 for svar med eksakte beløp og 11 for svar som var avrundet til nærmeste gangbare mynt.

## Koder til spørsmål 3

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 8,75 | 88 | 61,5 | 0,260 | 0,33 |
| 11 | 8,50 | 7 | 4,9 | 0,585 | 0,13 |
| | **Ikke riktig svar** | | | | |
| 70 | 9 | 6 | 4,2 | 0,340 | 0,07 |
| 71 | 11,25 | 12 | 8,4 | -0,114 | -0,04 |
| 79 | Andre svar | 22 | 15,4 | -0,865 | -0,37 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 8 | 5,6 | -1,080 | -0,26 |

Spørsmålet fikk en p-verdi på 0,65 som er mer i samsvar med hva en ønsker av en testoppgave, og betraktelig lavere enn for det foregående spørsmålet selv om spørsmålet bare krevde en liten ekstra utregning.

Riktig svar og spesielt kode 11 diskriminerer bra. Vi ser at kodene 10 og 11 har ”point-biserial” på henholdsvis 0,33 og 0,13. Slår vi i sammen begge kodene for riktig svar, vil spørsmålet korrelerer bedre med testen som helhet. Dette gir en ”point-biserial” på 0,40. Verdien viser at spørsmålet diskriminerer bedre enn det inntrykket vi får ved å se ”point-biserial” for hver av de riktige kodene.

De 7 besvarelsene som ble kodet 11 hadde betraktelig høyere gjennomsnittlig z-skåre enn besvarelsene med kode 10. De 7 hadde gitt svar etter den avrundingsregel som praktiseres. Siden elevenes mestring av avrundingsregelen ikke var gjenstand for test, valgte jeg å benytte kode 11. Det var heller ikke testens intensjon å prøve hvorvidt elevene mestret denne regelen siden spørsmålet senere skulle benyttes i PISA-undersøkelsen[13].

Hvis dette derimot hadde vært intensjonen, kunne det vært naturlig å sette kode 20 som gir 2 poeng, i stedet for kode 11. Jeg har i liten grad benyttet kode 20. I de åpne oppgavene er den blitt benyttet der det var ønske om forklaring eller utregning.

De seks elevbesvarelsene som er kodet 70, benyttet også avrundingsregelen. Her ble innkjøpet rundet nedover slik at de fikk igjen 50 øre for mye etter handelen. Det kan diskuteres om disse skulle vært kodet 12, men under tvil valgte jeg å kode dem under **Ikke riktig svar**. En svakhet med spørsmål 3 er at den ikke ble fulgt opp med utregning slik at man kunne kontrollere om feil oppfatning av regelen var årsaken til svaret. Det er dessverre mange som oppfatter regelen feil ved at både 25 øre og 75 øre blir avrundet nedover til

[13] Få av de land som er med i PISA-undersøkelsen praktiserer de samme avrundingsregel som i Norge. Andre land som f. eks. Danmark, runder av til nærmeste 25-øre. De fleste land praktiserer ingen slik regel. Derfor opererer PISA med et oppdiktet land som kalles *Zedland* hvor de bruker myntenheten *zed.* Dette ”landet” praktiserer ingen slike avrundingsregler.

nærmeste krone eller 50-øring. Problemet rundt avrundingsregelen kunne vært unngått dersom spørsmålet ble gitt med et kjøp på f. eks 4,8 kg poteter.

Kode 71 er forbeholdt de besvarelsene som regnet ut riktig pris for kjøpet, men oppga dette som riktig svar uten å se på differensen mellom kr 20 og den pris som ble betalt. Det er opplagt at de ikke leste spørsmålet godt nok. Dette er et problem for ikke bare svake elever. En frekvenstabell viser at også flere elever med høy z-skåre, blant annet en av de som gjorde det best på hele testen, ”slurvet” her.

Det var vaskelig å finne årsaken til de oppgitte svarene til de 22 elevene under kode 79, men 21 av disse hadde negativ z-skåre. Som ventet ble de svakeste elevbesvarelsene kodet 99 på dette spørsmålet med en gjennomsnittlig z-skåre på –1,080. Her viser også en frekvenstabell at alle hadde negativ z-skåre.

---

**Spørsmål 4:** POTETER

Vi kan sette opp et uttrykk som viser sammenhengen mellom x = antall *kg* og y = antall kroner hun har igjen.

Hvilket av disse utrykkene viser denne sammenhengen? Sett en ring rundt det svaret du mener er riktig.

A y = 2,50x – 20

B y = 20x – 2,50

C y = 20 – 2,50x

D y = 2,50 – 20x

---

Dette er testens første flervalgsoppgave, en oppgaveform norske elever har arbeidet lite med. Det dukket faktisk opp en flervalgsoppgave til avgangsprøven i grunnskolen våren 2001, bare noen uker etter at denne testen ble gjennomført. Der skulle elevene velge det funksjonsuttrykket blant andre alternativ som skulle passe til en rett linje tegnet i et koordinatsystem.

Grafer og funksjoner har en sentral plass i ungdomstrinnet, noe læreplanen understreker.

> *De skal kunne bruke sine kunnskaper om grafer og funksjoner til å undersøke og beskrive situasjoner og sammenhenger og til å arbeide med praktiske og matematiske problemer. (L97, s.166)*

For å løse denne oppgaven bør elevene være i besittelse av følgende kunnskaper og kunne kombinere disse:

- ha kjennskap om funksjonsuttrykk og kunne tolke dem.
- ha kjennskap til symbolspråket til variable størrelser i funksjonsuttrykket.
- kunne beskrive situasjoner i dagliglivet ved hjelp av funksjoner.

## Koder til spørsmål 4

Riktig svar er merket med *.

| **Koder** | **Frekvens** | **Prosent** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| A | 31 | 21,7 | -0,250 | -0,13 |
| B | 12 | 8,4 | -0,488 | -0,15 |
| C* | 81 | 56,6 | 0,377 | 0,44 |
| D | 4 | 2,8 | -0,961 | -0,16 |
| **Ikke svart** | | | | |
| 99 | 15 | 10,5 | -0,871 | -0,30 |

En svarprosent på nesten 90 anser jeg som meget tilfredsstillende for denne type oppgave tatt i betraktning at flere elever var lite kjent med flervalgsoppgaver.

Tabellen viser at 56,6 prosent av elevene valgte riktig alternativ, noe som ga en tilfredsstillende p-verdi på 0,57. Den gjennomsnittlige z-skåren er på 0,377 for riktig svar. Den er høy tatt i betraktning at så mange valgte riktig alternativ. Oppgavens "point-biserial" korrelasjon på riktig svar er 0,44 som viser at den har en god diskrimineringsevne.

Alle distraktorer har negativ gjennomsnittlig z-skåre, med en større tilslutning til A enn til de andre. Den store tilslutningen kan muligens komme av at de elevene som valgte denne hadde forståelse for symbolspråket til variable størrelser ved at pris per kg skal multipliseres med antall kg. Det viste at distraktoren fungerte utmerket. Den distraktoren som fikk dårligst tilslutning var D, og en gjennomsnittlig z-skåre på -0,961 viser at det var de aller svakeste som valgte dette alternativet.

Denne distraktoren var ikke velvalgt, noe den lave tilslutningen på bare 2,8 % viser. I følge de kriterier som blir lagt til grunn for analysen av pilotoppgaver, sies blant annet:

> *Distraktorene skal midlertidig være så "gode" at svarfrekvensen på hver av dem bør være minst 5 %. (Kleve 1994, s.9)*

Når det også bare var elever blant de aller svakeste som valgte dette alternativet, styrker dette at distraktor D enten bør omarbeides eller byttes ut.

En interessant side ved dette spørsmålet er å undersøke om jentene og guttene i samme grad ser sammenhengen mellom et funksjonsuttrykk og en situasjon i dagliglivet. I den følgende krysstabellen sammenlignes jentenes og guttenes besvarelse på spørsmål 4. Bare de som besvarte spørsmålet er tatt med.

**Krysstabell mellom kjønn og spørsmål 4**

Count

| | | Spørsmål 4 | | | | |
|---|---|---|---|---|---|---|
| | | A | B | C* | D | Total |
| Kjønn | Jenter | 15 | 2 | 53 | 1 | 71 |
| | Gutter | 16 | 10 | 28 | 3 | 57 |
| Total | | 31 | 12 | 81 | 4 | 128 |

Det fremkommer en klar sammenheng mellom kjønn og besvarelse. Nesten ¾ av jentene valgte riktig alternativ mens under halvparten av guttene gjorde det. En Kji-kvadrattest på krysstabellen gir verdi 8,87 og viser derved en signifikant sammenheng mellom kjønn og hvordan spørsmål 4 ble besvart.

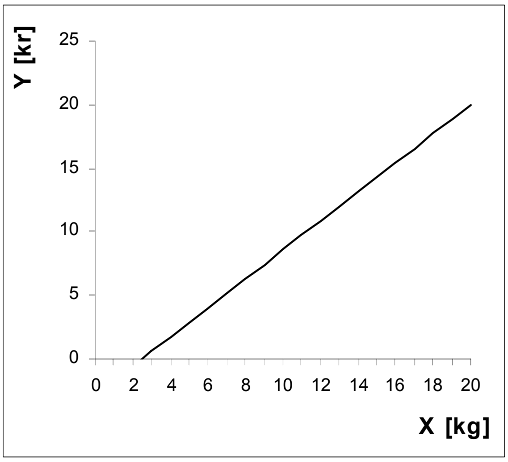
Forrige runde av PISA viste klart at jentenes gjennomsnittsskåre i lesing var bedre enn guttenes. Derfor bør oppgaveformuleringen ta hensyn til dette (se p.5.1). Ordbruken i denne oppgaven var kort og direkte. Forskjellen i besvarelsene kan derfor neppe ligge her. Min konklusjon er heller at jentene var flinkere til å se sammenhengen mellom oppgavens innledning og de alternative funksjonsuttrykkene.

**Spørsmål 5:** POTETER

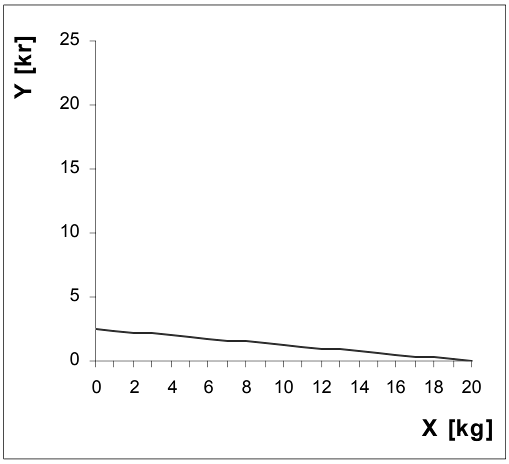
Hvilken av grafene nedenfor viser best sammenhengen mellom antall kroner hun har igjen (Y) og antall kg poteter som hun har kjøpt (X)?

Sett en ring rundt bokstaven ved den figuren du mener er riktig

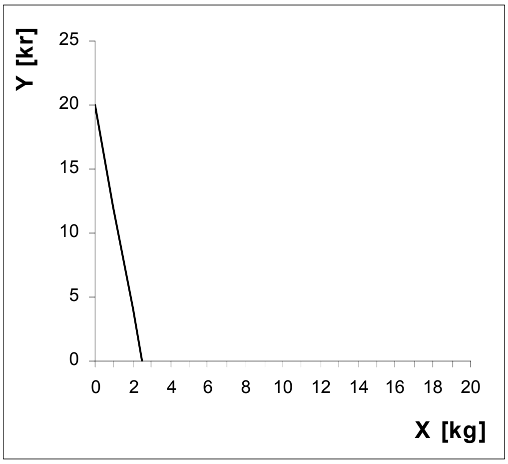
A



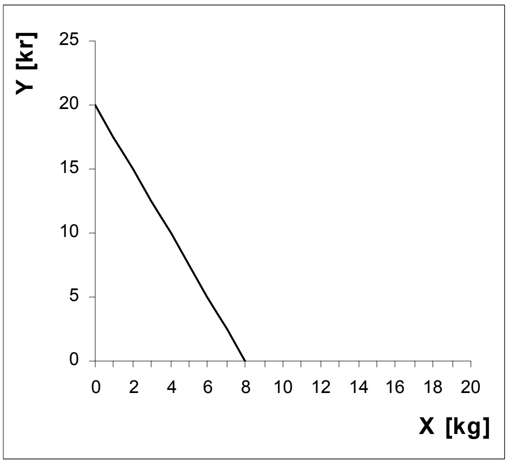

B




C




D




---

Dette spørsmålet er også en flervalgsoppgave, nær knyttet til det foregående spørsmålet. En forutsetning for å velge riktig alternativ på dette spørsmålet er at eleven må:

- ha kjennskap til koordinatsystemet.
- kunne undersøke grafers egenskaper.
- kunne utnytte kunnskapene om lineære funksjoner til å løse praktiske oppgaver.

# Koder til spørsmål 5

Riktig svar er merket med *.

| **Koder** | **Frekvens** | **Prosent** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| A | 28 | 19,6 | -0,286 | -0,14 |
| B | 2 | 1,4 | -0,725 | -0,09 |
| C | 12 | 8,4 | -0,705 | -0,21 |
| D* | 91 | 63,6 | 0,261 | 0,35 |
| **Ikke svart** | | | | |
| 99 | 10 | 7 | -0,583 | -0,16 |

Vi ser av tabellen at både svarprosenten og p-verdien også på denne flervalgsoppgaven er høy tatt i betraktning at også denne er vurdert å ligge i kompetanseklasse 2. Årsaken kan være at i motsetning til ved åpne oppgaver, er det enklere for en elev å krysse av et alternativ som synes rimelig i en flervalgsoppgave enn selv å prøve å regne på oppgaven. Det er tross alt hele 25 % sjanse for å tippe riktig. Den lave tilslutning til distraktorene (30,9 % av de som svarte på spørsmålet) viser at det her var liten grad av tipping. Dette styrkes ved at alle distraktorene som ventet fikk negativ gjennomsnittlig z-skåre.

Grunnen til at riktig alternativ ikke fikk høyere gjennomsnittlig z-skåre enn 0,261, skyldes det store antallet som valgte riktig, og at 47,3 % av disse hadde negativ z-skåre på hele testen. Vi ser også at disktraktor B burde omarbeides med samme begrunnelse som distraktor D i den første flervalgsoppgaven.

Slik både spørsmål 4 og spørsmål 5 er konstruert henger de på mange måter sammen. De representerer henholdsvis funksjonsuttrykket og grafen til samme problemstilling. Elevene skal vise forståelse for både lineære funksjoner og presentasjonen av disse i et koordinatsystem. (se L97, s.166)

Det var 56,6 % som krysset av riktig alternativ på spørsmål 4 og hele 63,6 % som krysset av riktig på spørsmål 5. Det er naturlig å se på hvordan de riktige alternativene fremkommer i en krysstabell for begge spørsmålene.

**Krysstabell for spørsmål 4 og spørsmål 5**

Count

| | | Spørsmål 5 | | | | | |
|---|---|---|---|---|---|---|---|
| | | A | B | C | D* | Blankt | Total |
| Spørsmål 4 | A | 6 | 2 | 3 | 20 | | 31 |
| | B | 2 | | 2 | 8 | | 12 |
| | C* | 16 | | 6 | 56 | 3 | 81 |
| | D | 3 | | | 1 | | 4 |
| | Blankt | 1 | | 1 | 6 | 7 | 15 |
| Total | | 28 | 2 | 12 | 91 | 10 | 143 |

.

Krysstabellen viser at 56, dvs. 39,2 %, av elevene krysset av riktig alternativ på begge flervalgsoppgavene. Det vil igjen tilsvare 44,8 % av de som krysset av på begge spørsmålene. Vi ser også at bare 7 elever ikke krysset av på noen av oppgavene og 51 elever valgte riktig alternativet på det ene spørsmålet og feil på det andre. 18 elever krysset av feil på begge. På tross av den store andelen som svarte riktig på begge spørsmålene, viser fordelingene av de 51 besvarelsene som har riktig svar på bare det ene spørsmålet og ikke det andre at det ikke er sterk sammenheng i svarmønsteret på begge spørsmålene. Dette viser at mange elever er usikre på funksjonsuttrykket og dets tilhørende graf, eller omvendt. En Kji-kvadrattest på de som besvarte begge spørsmålene, understreker dette. Det kan også legges til at begge oppgavene sett samlet bedret diskrimineringsevnen slik at ”Point-biserial” korrelasjonen nå ble 0,48, høyere enn for begge oppgavene sett hver for seg.

En tabell over svarkombinasjonene inkludert z-skårer for de som svarte riktig på minst ett av spørsmålene, ser slik ut:

## Kombinasjonstabell for spørsmål 4 og spørsmål 5 for besvarelser med minst ett riktig svar

| **Spørsmål 4** | **Spørsmål 5** | **Frekvens** | **Prosent** | **Z-skåre** |
|---|---|---|---|---|
| C* | D* | 56 | 39,2 | 0,593 |
| A | D* | 20 | 14 | -0,003 |
| C* | A | 16 | 11,2 | -0,001 |
| C* | C | 6 | 4,2 | -0,469 |
| B | D* | 8 | 5,6 | -0,355 |
| D | D* | 1 | 0,7 | -1,139 |
| C* | Blankt | 3 | 2,1 | 0,044 |
| Blankt | D* | 6 | 4,2 | -0,903 |

Tabellen viser ved en gjennomsnittlig z-skåre på hele 0,593 at det i hovedsak var de som skåret best på testen, som traff riktig på begge spørsmålene. En frekvenstabell viser at 37 av de 56 besvarelsene hadde positiv z-skåre. Av de 25 elevene som gjorde det best på testen, var det 21 som svarte riktig på begge spørsmålene.

Det var 65 elever med negativ z-skåre som besvarte begge spørsmålene. Av disse var det 19 som krysset av riktig på begge. Det er vanskelig å fastslå hvor mange som krysset av etter ”overbevisning” eller tipping. Sannsynligheten for å tippe riktig på begge er 6,25 %, mens vi her har over 29 % som svarte riktig. Dette kan tolkes som at mange hadde en klar oppfatning av hvordan funksjonsuttrykket og grafen til oppgaveteksten ville se ut. Av de tre som svarte riktig på spørsmål 4, men ikke svarte på spørsmål 5, var det to med positiv

z-skåre. Mer bemerkelsesverdig er det at de 6 som unngikk å svare på spørsmål 4, men svarte riktig på spørsmål 5, alle hadde en sterk negativ z-skåre med gjennomsnitt på –0,903. Det kan kanskje tyde på at noen tippet riktig alternativ. Vi ser også at 51 elever svarte riktig på ett av spørsmålene, men feil på det andre. Av disse var det 31 med negativ z-skåre og 20 med positiv.

### 7.1.1 Oppsummering rundt oppgaven Poteter

Under utvikling av denne oppgaven hadde jeg blant annet i tankene å starte med en oppgave som også de aller svakeste kunne prøve seg på. Svarprosenten på disse spørsmålene viste at jeg lyktes med intensjonen. Er det også mulig å kombinere denne intensjonen med oppgaver som kan forsvares som gode testoppgaver? Det kan diskuteres.

De fleste som har undervist i matematikk noen år, har observert en gruppe elever med emosjonelle problemer knyttet til matematikk og da spesielt når de blir prøvd i faget i forbindelse med prøver og tester. Mange av disse elevene forsøker å trekke seg ut av situasjonen og viser derfor ikke hva de egentlig mestrer.

Snorre Ostad uttrykker det slik:

> *Emosjonelle faktorer spiller naturligvis en sentral rolle for prestasjonene i matematikk – som i andre fag. Hos påfallende mange matematikksvake elever finner en symptomer på angst, utrygghet og usikkerhet: De vegrer seg for å starte arbeidet, forsøker ofte å trekke seg tilbake fra arbeidssituasjonen etc. (Ostad 1993, s.25)*

I en test der matematikksvake elever vegrer seg for å vise sine kunnskaper eller mangel på sådanne, er det fare for at validiteten blir redusert, spesielt hvis elevgruppen har lav motivasjon for å gjennomføre kunnskapstesten. Under gjennomføringen av testen ble det observert at der læreren gikk rundt og oppmuntret elevene til å fortsette, var det tydelig at flere forsøkte seg på nye oppgaver. I andre grupper der læreren forholdt seg i ro og elevene ikke fikk den stimuleringen, anstrengte de seg i mindre grad. Derfor måtte prestasjonene for gruppene bli forskjellige. Intensjonen med uttestingen var å se om oppgavene egnet seg som testoppgaver og ikke hvordan de enkelte gruppene presterte i forhold til hverandre. Derfor lot jeg faglærerne styre denne prosessen.

Det var oppmuntrende å se at alle elevene besvarte minst ett spørsmål i oppgaven. Det kan igjen tyde på at oppgaver med kontekst lokalisert ”nær” elevene bidrar til å høyne svarprosenten.

Av de 119 elevene som svarte på alle spørsmålene i oppgaven, var det 45 elever som også svarte riktig på alle spørsmålene. Noe overraskende var det at 10 av disse hadde negativ z-skåre på hele testen. Dette viser at oppgaven i helhet passet for også svake elever. På tross av dette diskriminerte hele oppgaven tilfredsstillende med ”point-biserial” på 0,57. Den gjennomsnittlige z-skåren for de som svarte riktig på alle spørsmålene, ble naturligvis betraktelig høyere, 0,840, enn for hvert av spørsmålene.

Dette var den eneste oppgaven i pilottesten med negativ korrelasjon mellom to av spørsmålene i samme oppgaveenhet. Av de fem som ikke oppga riktig svar på spørsmål 2, alle med sterk negativ z-skåre, valgte tre riktig alternativ på spørsmål 4. Dette var tilstrekkelig til at disse to spørsmålene korrelerte negativt med hverandre.

Denne oppgaven ble ikke med på Generalprøven i PISA, siden den hadde svake sider. Flere av spørsmålene hadde for høy p-verdi. For å sikre validiteten til et spørsmål er det viktig at det har en rimelig vanskegrad slik at de flinkeste elevene får vist hva de kan. Styrken ved oppgaven var etter min mening, de to flervalgsoppgavene som presenterte både funksjonsuttrykket og grafen til samme problemstilling, en oppgaveform Generalprøven ville tjent på å ha flere av. Elevene vil ved slike kombinasjoner av oppgaver kunne vise sine begrepsforståelser rundt relasjonen mellom et funksjonsuttrykk og tilhørende graf. Jeg anser at disse to spørsmålene sammen med konteksten til oppgaven har høy innholdsvaliditet både i forhold til L97 (se p.2.1 om Grafer og funksjoner) og PISA. I Guidelines er dette formulert slik:

> *Patterns, relations, functions; functions as formulas, and as graphs; relation between different classes of functions and their graphical representation; Algebra as a prerequisite for calculus. (OECD/PISA 2001a s.23)*

## 7.2 Oppgaven Trappekonstruksjon

Denne oppgaven er en videreutvikling av en eksamensoppgave for privatister jeg laget våren 1991 på oppdrag fra Rådet for videregående opplæring. Den gang var dette en trigonometrioppgave der elevene skulle beregne vinkler og avstander på en trapp.

I spørsmål 6 blir elevene prøvd i en oppgave som tar utgangspunkt i en trappekonstruksjon med angitte mål for høyde, og der antall trinn er tegnet inn. Ved å betrakte figuren og bruke de informasjoner som der er gitt, skal elevene beregne trinnhøyden for den gitte figuren. Derfor ligger spørsmålet innenfor *Rom og form*.

Selv om *Rom og form* i følge L97 er et målområde for elevene de første årene i grunnskolen, vil elevene også senere stifte bekjentskap med dette gjennom konstruksjoner og romfigurer. Dette målområdet skal senere i grunnskolen være med på å utvikle begreper i geometri gjennom å klassifisere figurer etter deres egenskaper.

De andre spørsmålene i oppgaven tar alle utgangspunkt i *Forandring og sammenheng*, og utfordringen blir gitt ved presentasjon av en *Trappeformel* som gjelder for konstruksjon av trapper. Elevene skal her benytte denne formelen for å regne seg fram til trappens dimensjoner med de gitte forutsetninger.

Konteksten for denne oppgaven er tatt fra en autentisk situasjon og vil naturlig plasseres i *Arbeidslivet*. Både spørsmål 6 og spørsmål 7 har jeg vurdert til å ligge i kompetanseklasse 1, spørsmål 8 og spørsmål 9 i kompetanseklasse 2, og spørsmål 10 i kompetanseklasse 3.

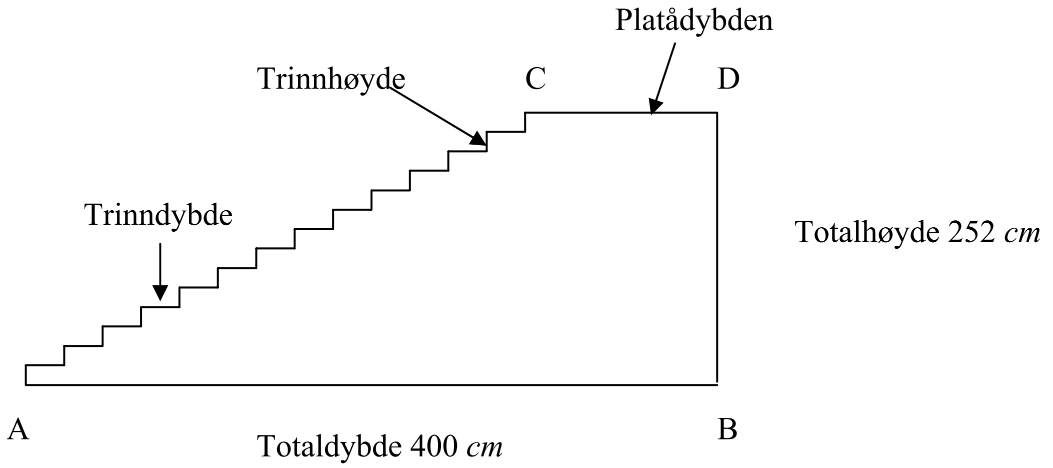
# TRAPPEKONSTRUKSJON

Påsken er en tid hvor det ofte skjer bruddskader. Det er ikke bare i skisporet dette skjer. Mange hyttetrapper er bygget av ”amatører” som ikke har tatt hensyn til trappeformelen. Man vil faktisk snuble oftere i trapper som ikke er konstruert etter denne formelen.

I et utsnitt fra Byggforskningsserien A 532.111 er trappeformelen forklart:

> Trinndybden er avhengig av trinnhøyden og bør nærme seg skrittlengden. For å bestemme trinndybden eller trinnhøyden kan man bruke ”trappeformelen”:
>
> **2 trinnhøyder + 1 trinndybde = 63 *cm*.**
>
> Eksempel: Hvis trinnhøyden er 15 cm, blir trinndybden 63 *cm* – 2 x 15 *cm* = 33 *cm*.

Figuren nedenfor viser en trapp med 14 trinn og en høyde på 252 *cm*:




**Spørsmål 6:** TRAPPEKONSTRUKSJON

Vis ved utregning at trinnhøyden blir 18 *cm*.

Det forventes følgende forutsetninger for å kunne svare på dette spørsmålet:

- Eleven behersker bruken av de fire regningsarter.
- Eleven kan klassifisere figurer og undersøke deres mønster.
- Eleven kan registrere, formulere og arbeide med problemer knyttet til arbeidsliv.

## Koder til spørsmål 6

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 252cm : 14 =18cm | 73 | 51 | 0,483 | 0,50 |
| 11 | 252cm : 18cm =14 | 1 | 0,7 | -0,429 | -0,04 |
| | **Ikke riktig svar** | | | | |
| 70 | 63cm -2 *x* 18cm = 27cm | 14 | 9,8 | 0,002 | 0,00 |
| 79 | Andre svar | 9 | 6,3 | -0,022 | -0,01 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 46 | 32,2 | -0,753 | -0,52 |

Det åpenbart riktige svaret er kodet 10, men jeg valgte også å godta svaret gitt kode 11, som riktig. Samlet ga kodene 10 og 11 en p-verdi på 0,52. Spørsmålet diskriminerte tilfredsstillende med "point-biserial" på 0,489 for de to riktige kodene sett under ett. Kandidaten på kode 11 ga et noe uventet svar ved å dividere trappens høyde med trinnhøyden slik at det ga 14 trinn som både trappefiguren og teksten opplyste. Ikke uriktig selv om en annen framgangsmåte for å svare på spørsmålet var forventet. Dette var en "svak" elev, noe z-skåren også viser.

De 14 besvarelsene som er kodet 70, benyttet trappeformelen for å bekrefte trinnhøyden uten å innhente informasjon fra tekst og figur. Det er sannsynlig at de fleste her gikk i kast med spørsmålet etter at de hadde løst neste spørsmål og benyttet informasjoner de der fikk gjennom svaret. Svarene under denne kode hadde en gjennomsnittlig z-skåre like over null.

Det er en tankevekker at hele 46 elever, nesten 1/3, ikke svarte på det første spørsmålet i en oppgave det var naturlig å sette i kompetanseklasse 1. Oppgaven krevde bare at elevene dividerte trappehøyden med antall trinn. Det viser at mange elever har lite trening i å hente ut relevante informasjoner fra en tekst knyttet til en illustrasjon. Oppsettet av oppgaven i sin helhet må muligens også ta sin del av skylden. Et annet oppsett der trappeformelen ble presentert først etter spørsmålet, kunne trolig ført til at flere ville besvart spørsmålet (se p. 7.2.1).

**Spørsmål 7:** TRAPPEKONSTRUKSJON

Hva blir trinndybden når trinnhøyden er 18 *cm*. Vi antar her at trappeformelen skal gjelde?

……..*cm*

## Koder til spørsmål 7

| Kod | Svaralternativer | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 27cm | 73 | 51 | 0,535 | 0,55 |
| | **Ikke riktig svar** | | | | |
| 70 | 25cm – 26cm | 3 | 2,1 | -0,824 | -0,12 |
| 71 | 28cm – 29cm | 10 | 7 | 0,340 | 0,09 |
| 79 | Andre svar | 11 | 7,7 | -0,655 | -0,19 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 46 | 32,2 | -0,712 | -0,49 |

Det er gitt én kode, 10, for riktig svar, noe som er naturlig for dette spørsmålet som krever ett svar uten å vise til utregning eller forklaring. Denne koden hadde også en litt høyere gjennomsnittlig z-skåre enn på spørsmål 6, selv om p-verdien var den samme, altså 0,51. Det kan skyldes at spørsmålet passet noe bedre for de "flinkere" elevene. Selv om spørsmålet matematisk sett trolig ga en større utfordring enn spørsmål 6, var det naturlig å sette det i kompetanseklasse 1. Vi ser også at spørsmålet diskriminerte bra med "point-biserial" på 0,55.

Det var tre koder under **Ikke riktig svar**, og av disse hadde kode 71 en relativ høy gjennomsnittlig z-skåre, noe som tyder på at flere "flinke" elever resonerte seg til et svar. Under denne koden var svaralternativet fra *28cm til og med 29cm*. Ser vi nærmere på en frekvenstabell for z-skåren får vi en forklaring på den relativt høye gjennomsnittlige z-skåren. Det var seks elever med positiv z-skåre. Hvordan de kom fram til disse svarene er det vanskelig å fastslå siden spørsmålet ikke gjorde krav på å vise utregning eller forklaring. Flere oppgaver som ikke krever at elevene viser forklaring, kan gjerne erstattes med flervalgsoppgaver der elevene merker av riktig alternativ, eller de kan som i dette tilfellet, sette inn de respektive alternativene for å se hvilket som vil passe inn i trappeformelen.

**Krysstabell mellom spørsmål 6 og spørsmål 7**

Count

| | | Spørsmål 7 | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 10 | 70 | 71 | 79 | 99 | |
| Spørsmål 6 | 10 | 56 | 2 | 7 | 5 | 3 | 73 |
| | 11 | 0 | 0 | 0 | 0 | 1 | 1 |
| | 70 | 12 | 0 | 0 | 0 | 2 | 14 |
| | 79 | 1 | 0 | 2 | 2 | 4 | 9 |
| | 99 | 4 | 1 | 1 | 4 | 36 | 46 |
| Total | | 73 | 3 | 10 | 11 | 46 | 143 |

En krysstabell for spørsmål 6 og spørsmål 7 viser svarmønstret for de to spørsmålene. På begge spørsmålene var det 46 svar som ble kodet 99, altså ikke besvart. Det er verdt å bemerke at så mange som 36 elever, over 25 %, ikke besvarte noen av spørsmålene. Dette er betenkelig når man tar i betraktning at i begge spørsmålene skulle elevene anvende rutinemessige utregninger. Spørsmålene var derfor plassert i kompetanseklasse 1, fordi det åpenbart er mindre kompleksitet i kompetanseklasse 1 enn i de andre kompetanseklassene. Problemet kan trolig ligge i den innledende konteksten som for mange elever kan ha vært for omfattende og uoversiktlig. De ble ikke tilstrekkelig tent på å i det minste prøve å svare på spørsmålene. For mange elever er det viktig hvordan en matematikkoppgave blir presentert. I denne oppgaven var det antagelig uheldig at både trappeformelen og trappekonstruksjonen ble presentert samtidig. (se oppsummeringen p.7.2.1)

Krysstabellen viser også at av 14 besvarelsene som ble kodet 70 i spørsmål 6, var det 12 som ble kodet 10 i spørsmål 7. Slik spørsmål 7 var formulert og slik de samme 12 besvarte spørsmål 6, er det trolig at de benyttet svaret i det ene spørsmålet for senere å gå tilbake til det foregående og bruke det svaret i sin besvarelse. Besvarelsen ble ikke riktig i forhold til spørsmålet, men de viste hvordan trappeformelen kunne brukes.

Man ser også at 56 elever svarte riktig på begge spørsmålene. Av disse var det 18 elever med negativ z-skåre på testen og 38 med positiv. Den gjennomsnittlige z-skåre ble på 0,633.

---

**Spørsmål 8:** TRAPPEKONSTRUKSJON

Sett en ring rundt Ja eller Nei for hvert av de følgende utsagnene om en trapp laget etter trappeformelen:

| **Utsagn:** | **Mulig?** |
|---|---|
| Man kan forandre trinnhøyden uten samtidig å forandre trinndybden. | Ja / Nei |
| Det er mulig å lage en trapp der både trinndybde og trinnhøyde er 20 *cm*. | Ja / Nei |
| Trappen vil bli brattere dersom vi øker trinndybden | Ja / Nei |
| Det er mulig å lage en trapp der trappetrinnet har en høyde som er lik trinndybden. | Ja / Nei |

---

## Koder til spørsmål 8

| Koder | Frekvens | Prosent | Prosent av de som svarte | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 1 | 36 | 25,2 | 28,3 | 0,842 | 0,49 |
| | **Ikke riktig svar** | | | | |
| 0 | 91 | 63,6 | 71,7 | -0,160 | -0,21 |
| | **Ikke svart** | | | | |
| 99 | 16 | 11,2 | | -0,984 | -0,35 |

I en sammensatt flervalgsoppgave har eleven to alternativer for hvert spørsmål, ja eller nei. Ved ren tipping gir det 50 % sjanse for å tippe riktig på ett av utsagnene eller 6,25 % på alle fire utsagnene. På spørsmål 8 var det 25,2 % som svarte riktig på alle utsagnene, eller 28,3 % av de som svarte på oppgaven. Med en svarprosent på 88,2 % vil det være naturlig å forvente en p-verdi på langt over 0,5 for de enkelte utsagn. Kode 99 representerer antallet som ikke krysset av på noen av utsagnene. Oppgaven diskriminerte bra mellom de flinkeste og de andre med "point-biserial" på 0,49 for riktig svar. En frekvenstabell viser at 28 elever under kode 1 hadde positiv z-skåre. Både **Ikke riktig** og **Ikke svart** hadde tilfredsstillende negative verdier for både gjennomsnittsskåren og "point-biserial". Det var dessuten positiv korrelasjon mellom riktig svar på alle utsagnene. Vi ser også at p-verdien ble på 0,25.

## Koder til de respektive utsagn i spørsmål 8

Tabellen for de enkelte utsagnene blir følgende der de riktige svarene er merket *:

| **Koder til de respektive utsagn** | **Frekvens** | **Prosent** | **Prosent av de som svarte** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|---|
| **Utsagn A** | | | | | |
| Ja | 24 | 16,8 | 19 | -0,331 | -0,15 |
| Nei* | 102 | 71,2 | 81 | 0,241 | 0,38 |
| 99 | 17 | 11,9 | | -0,979 | -0,36 |
| **Utsagn B** | | | | | |
| Ja | 34 | 23,8 | 27,4 | -0,384 | -0,22 |
| Nei* | 90 | 62,9 | 72,6 | 0,307 | 0,40 |
| 99 | 19 | 13,3 | | -0,766 | -0,30 |
| **Utsagn C** | | | | | |
| Ja | 48 | 33,6 | 38,1 | -0,249 | -0,20 |
| Nei* | 78 | 54,4 | 61,9 | 0,367 | 0,40 |
| 99 | 17 | 11,9 | | -0,882 | -0,33 |
| **Utsagn D** | | | | | |
| Ja* | 81 | 56,6 | 66,9 | 0,288 | 0,33 |
| Nei | 40 | 28 | 33,1 | -0,249 | -0,16 |
| 99 | 22 | 15,4 | | -0,607 | -0,26 |

Her er både den gjennomsnittlige z-skåren og ”point-biserial” lavere for hvert av utsagnene i forhold til spørsmålet sett under ett. Dette er naturlig siden det er betydelig lettere å treffe riktig på ett utsagn enn å treffe riktig på alle fire.

Hvis man ser bort fra de som ikke svarte, vil en krysstabell mellom de forskjellige utsagnene ikke vise noe klart mønster med unntak av for 8C og 8D.

Ser vi på krysstabellen mellom 8C og 8D får vi følgende:

**Krysstabell mellom utsagnene C og D**

Count

| | | Utsagn D | | | Total |
|---|---|---|---|---|---|
| | | 0 | 1 | 99 | |
| Utsagn C | 0 | 21 | 23 | 4 | 48 |
| | 1 | 19 | 58 | 1 | 78 |
| | 99 | 0 | 0 | 17 | 17 |
| Total | | 40 | 81 | 22 | 143 |

Det var totalt 121 elever som krysset av på begge utsagnene, og av disse var det 58 som krysset av riktig og 21 som krysset av feil på begge.

I motsetning til de andre kombinasjonene var det her i større grad de samme elevene som krysset av riktig på det ene utsagnet som også krysset av riktig på det andre. Tilsvarende for de som krysset av feil. En Kji-kvadrattest har på 5 % nivå en kritisk verdi på 3,84 mens testen her ga en verdi langt over, på hele 6,72. Den ligger faktisk over den kritiske verdien på 1 % nivå. Derfor kan vi slå fast at det er en sterkere sammenheng mellom disse to utsagnene enn på de andre kombinasjonene. Det kan da være naturlig å forandre eller droppe ett av disse utsagnene.

Denne flervalgsoppgaven ble også gitt under Generalprøven, uten spørsmål 8D. Analysen som her er foretatt viser da også at dersom ett av utsagnene skulle sløyfes måtte det være enten 8C eller 8D på grunn av den sterke sammenhengen mellom disse to utsagn.

Krysstabellen mellom utsagnene 8A og 8D gir et interessant mønster.

**Krysstabell mellom utsagnene A og D**

Count

| | | Utsagn D | | | Total |
|---|---|---|---|---|---|
| | | 0 | 1 | 99 | |
| Utsagn A | 0 | 8 | 16 | 0 | 24 |
| | 1 | 32 | 64 | 6 | 102 |
| | 99 | 0 | 1 | 16 | 17 |
| Total | | 40 | 81 | 22 | 143 |

Her var det også 121 elever som krysset av på begge utsagnene. Vi ser at det var flere som krysset av riktig og færre som krysset av feil på begge utsagnene i sammenligning med forrige krysstabell. Vi ser at 2/3 av de som krysset av feil på utsagn 8A, krysser av riktig på utsagn 8D. Det samme forholdet gjelder også for de som krysset av riktig på det første utsagnet. En Kji-kvadrattest gir under slike forhold verdien null. Det kan igjen tolkes som at det er svært liten sammenheng mellom besvarelsene på disse utsagnene.

**Spørsmål 9:** TRAPPEKONSTRUKSJON

Det skal lages en trapp der totalhøyden er 252 *cm* (BD) og totaldybden (AB) er 400 *cm*.

Hvor mange trinn vil en trapp få hvis trappedybden er 29,4 *cm*?

…………trinn.

---

Her måtte elevene forholde seg til flere informasjoner der en av disse, totaldybden for hele trappen, ikke var relevant i forhold til spørsmålet. Denne opplysningen forvirret antagelig enkelte elever og kan ha forårsaket at noen kom "skjevt" ut i sine utregninger eller hoppet over spørsmålet[14]. En intensjon med spørsmålet var nettopp at elevene skulle hente ut de relevante informasjonene for å løse oppgaven ("retrieve information"). Selvfølgelig kunne spørsmålet presenteres kun med relevante informasjoner med trolig høyere p-verdi som resultat, men det ville måle noe annet enn det som nå var intensjonen.

Elevene måtte først finne trinnhøyden ved hjelp av trappeformelen og deretter dividere

totalhøyden for trappa med trinnhøyden for på den måten å komme til 15 trinn. Det er naturlig å plassere spørsmålet i kompetanseklasse 2.

Svarfordelingen med koder ble følgende:

## Koder til spørsmål 9

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 15 | 18 | 12,6 | 1,076 | 0,41 |
| | **Ikke riktig svar** | | | | |
| 70 | 14,16,17 | 18 | 12,6 | 0,484 | 0,18 |
| 71 | 12,13,18 | 20 | 14 | 0,434 | 0,18 |
| 79 | Andre svar | 22 | 15,4 | -0,370 | -0,16 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 65 | 45,5 | -0,440 | -0,40 |

[14] Flere elever ble muligens også forvirret av at trinndybden her ble kalt trappedybde.

65 elever, hele 45,5 %, besvare ikke spørsmålet. Trolig ville denne prosenten vært noe lavere hvis spørsmålet ble presentert med bare relevante opplysninger. Med en gjennomsnittlig z-skåre på –0,44 for de som ikke besvarte spørsmålet, er det grunn til å anta at de svakeste styrte unna. Under **Ikke riktig svar** var både kode 70 og 71 representert med relativt høy gjennomsnittlig z-skåre på henholdsvis 0,484 og 0,434. Frekvenstabeller for z-skåren for de besvarelsene som ble kodet 70 og 71, viser at mange besvarelser fra elever med høy z-skåre ble kodet der. ”Point-biserial” viser også at spørsmålet diskriminerte bra mellom de som ble kodet 70 og 71 i forhold til de som ble kodet 79 og 99. Dette viser at spørsmålet burde vært presentert som langsvaroppgave med utregning for å se hvordan de enkelte forholdt seg til de informasjoner som ble gitt i teksten. Det ville da gitt grunnlag til å kode 20 for helt riktig svar og kode 10 for delvis riktig.

Kode 10 viser at bare 18 elever oppnådde riktig svar. En p-verdi på bare 0,13 forteller at testteoretisk er denne oppgaven lite ønskelig. En gjennomsnittlig totalskåre for de som svarte riktig på hele 1,076 og ”point-biserial” på 0,41, skulle tilsi at oppgaven ble løst av de aller flinkeste elevene. Men en frekvenstabell viser også at fire av elevene hadde negativ z-skåre på hele testen. Blant disse var det én som svarte riktig på fire av spørsmålene i oppgaven, noe som skulle tyde på at oppgaven *Trappekonstruksjon* falt i ”smak”.

---

**Spørsmål 10:** TRAPPEKONSTRUKSJON

Hva blir platådybden (CD) dersom en trapp skal ha 12 trinn (totalhøyden skal fortsatt være 252 *cm* og totaldybden 400 *cm*)?

………*cm*

---

Spørsmålet elevene her gikk i kast med, oppfatter jeg som testens mest utfordrende. Det kan være naturlig å sette det i kompetanseklasse 3, selv om rammeverket for PISA uttrykker at en større problematisering bør være til stede for å kunne høre til i denne klassen.

På spørsmål 10 måtte elevene forholde seg til 4 relevante informasjoner inkludert trappeformelen. For å løse oppgaven måtte elevene først finne trappehøyden ved å dividere totalhøyden med trappens 12 trinn. Ved hjelp av trappeformelen regner de så ut trinndybden for trinnet. Det ble da elevens tolkning om det øverste trinnet inkluderes i platådybden. Før oppgaven ble rettet, ble begge alternativene akseptert som riktig svar, enten 148 cm eller 169 cm. Siden ingen ga svaret 169 cm, ble 148 cm det eneste riktige svar og derfor kodet 10.

## Koder til spørsmål 10

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 148 (169) cm | 6 | 4,2 | 1,720 | 0,36 |
| | **Ikke riktig svar** | | | | |
| 70 | 47,2 cm | 11 | 7,7 | 0,635 | 0,18 |
| 71 | 21 cm | 7 | 4,9 | 0,720 | 0,16 |
| 79 | Andre svar | 19 | 13,3 | -0,130 | -0,05 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 100 | 69,9 | -0,199 | -0,30 |

En p-verdi på 0,04 bekrefter at spørsmålet i stor grad krevde metoder og kombinasjoner på noe høyere matematisk nivå enn de andre spørsmålene i testen. De 6 som oppga det riktige svaret hadde da også en svært høy gjennomsnittlig z-skåre på 1,72. Selv om spørsmålet fikk en tilfredsstillende "point-biserial" for riktig svar, tatt i betraktning den lave p-verdien, diskriminerte den bra mellom de aller beste og de elevene som ble kodet under 70 og 71. Vi ser også av tabellen at både de under kode 70 og 71 også har høy gjennomsnittlig z-skåre.

De 11 besvarelsene i kode 70 hadde muligens misforstått spørsmål 10 ved at de tok med seg trappedybden 29,4 cm fra foregående spørsmålet og multiplisert med de 12 trinnene. Da har de tolket mer inn i oppgaven enn teksten gir grunnlag for, og trappeformelen ville heller ikke gjelde.

Kandidatene under kode 71 dividerte trappehøyden med 12 trinn og fikk en trinnhøyde på 21 cm. Dette vil gi, ifølge trappeformelen, en trinndybde på 21 cm som også var deres svar. Det er umulig å vurdere hvorvidt de misforstod spørsmålet eller om de fant en lettvint løsning på en komplisert oppgave. De var kanskje mer opptatt av å kunne gi et svar selv om det ikke nødvendigvis var riktig platådybde. De 43 som oppga et svar på dette spørsmålet, hadde en gjennomsnittlig z-skåre på hele 0,462.

På tross av den lave p-verdien diskriminerte spørsmålet tilfredsstillende med en "point-biserial" på 0,36 for kode 10. Det er opplagt at et spørsmål der nesten 70 % ikke svarer, ikke bør forekomme så tidlig i en test, om det i det hele tatt bør forekomme. Det er en fare for at det fratar enkelte motivasjonen for å fortsette testen.

**Krysstabell mellom spørsmål 9 og spørsmål 10**

Count

| | | Spørsmål 10 | | | | | |
|---|---|---|---|---|---|---|---|
| | | 10 | 70 | 71 | 79 | 99 | Total |
| Spørsmål 9 | 10 | 5 | 1 | 2 | 3 | 7 | 18 |
| | 70 | 0 | 2 | 3 | 4 | 9 | 18 |
| | 71 | 1 | 7 | 0 | 3 | 9 | 20 |
| | 79 | 0 | 1 | 2 | 9 | 10 | 22 |
| | 99 | 0 | 0 | 0 | 0 | 65 | 65 |
| Total | | 6 | 11 | 7 | 19 | 100 | 143 |

Krysstabellen for spørsmål 9 og 10 viser en sterk sammenheng mellom elevbesvarelsene på begge spørsmålene for de som svarte riktig og de som ikke svarte riktig henholdsvis på det ene spørsmålet eller begge. Kji-kvadratverdien mellom de som svarte riktig på begge spørsmålene, bare det ene eller feil på begge, ble da også så høy som 12,22. De fem elevene som svarte riktig på begge spørsmålene fikk en gjennomsnittlig z-skåre på hele 1,960.

### 7.2.1 Oppsummering rundt oppgaven Trappekonstruksjon

Trappeoppgaven har et stort potensial som testoppgave. Alle spørsmålene sirkulerte rundt trappeformelen og en trappekonstruksjon. Valg av problemstillinger rundt dette kunne være andre og kanskje bedre. Antall spørsmål der elevene og ikke oppgavene blir testet, kunne med fordel vært redusert. De åpne spørsmålene som det ikke kreves utregning eller forklaring til, kunne gjøres om til flervalgsoppgaver.

Oppgaven ble med noen forandringer tatt med i Generalprøven. Det er forståelig at spørsmål 10 (svært lav p-verdi) ble fjernet. Validiteten til en test blir svekket dersom flere oppgaver gjennomgående har for høy vanskegrad. Det er ikke bare de svakeste elevene som ikke får vist hva de faktisk kan ved slike oppgaver. Jeg vil samtidig presisere at det ikke er testen selv som blir validert, men hvilke konklusjoner vi kan trekke ut fra elevenes prestasjoner på en slik test.

I tillegg til at spørsmål 10 ble fjernet, ble det foretatt noen forandringer i de øvrige spørsmålene og presentasjonen til Generalprøven. Trappeformelen ble ikke presentert før elevene måtte gjøre bruk av den, for dermed å forhindre misforståelser rundt spørsmål 6. Formuleringen av dette spørsmålet ble forandret til *"Hva er høyden på hvert av de 14 trinnene?"* istedenfor *"Vis ved utregning at trinnhøyden blir 18 cm."* Når man tar i betraktning at så mange som 56 elever med positiv z-skåre oppfattet spørsmålet feil eller ikke besvarte spørsmålet, viser dette hvor viktig det er at det jobbes aktivt med spørsmålsformuleringen ved utvikling av oppgaver.

> *Remember, the criterion for determining a correct answer to a test item is that a consensus of knowledgeable persons could be reached about the best response. (Osterlind 1992, side 221 – 222)*

I tillegg til at det siste utsagnet på spørsmål 8 ble fjernet ble spørsmålet også omgjort fra at elevene skulle sette ring rundt **Ja** eller **Nei** til påstander der eleven setter ring rundt **Riktig** eller **Galt**.

> *As simple and direct measures of attainment, the true-false format can be used to cover a wide range of content. One can scarcely imagine a subject in which this format would be wholly inappropriate. True-false items can be prepared to cover factual knowledge as well as judgments and opinions; and, they can be used to assess knowledge from the very low level of simple recall to the more sophisticated cognitive processing levels. (ibid, side 226 – 227)*

Det var 13 elever som ikke besvarte noen av spørsmålene i oppgaven, mens 33 elever svarte på alle spørsmålene. Den gjennomsnittlige z-skåren ble henholdsvis -1,012 og 0,897. Dette viser tydelig at de aller svakeste elevene ikke presenterte noen svar på oppgaven, mens den flinkeste gruppen forsøkte seg på alle spørsmålene.

Som følge av den høye vanskegraden på spørsmål 10 var det bare fire elever som svarte riktig på alle fem spørsmålene i oppgaven. Disse var testens flinkeste med gjennomsnittlig z-skåre på hele 1,996. På tross av de få som svarte riktig på alle fem spørsmålene, ble "point-biserial" relativt høy, på hele 0,34. Oppgaven diskriminerte derfor bra mellom de aller flinkeste og de andre. Ser man på de fire andre spørsmålene som i noe omarbeidet form var med i Generalprøven, var det 10 elever med gjennomsnittlig z-skåre på 1,463 som svarte riktig på alle spørsmålene. Spredningen mellom disse var stor. En av disse elevene hadde en z-skåre på -0,193 på hele testen.

L97 legger mye vekt på at elevene utvikler begreper rundt *Forandring og sammenheng*:

> *Det må skapes en økt oppmerksomhet om selve variabelbegrepet og om hva formler og uttrykk kan tjene til. Temaet krever spesiell oppmerksomhet fordi det i noen grad bryter med tidligere tenkemåter. Den formelle siden ved algebraen må ha et grunnlag i arbeid med konkrete eksempler. Algebra blir et redskap til å løse problemer, et språk som kan lette tenkning og resonnement, og en kilde til å oppdage nye sammenheng. (L97, s.156)*

Spørsmål 8 mener jeg har høy validitet i forhold til intensjonene ved både dette og det femte felles mål i L97 (se p.2.1). Validiteten er også høy i forhold til PISAs rammeverk. Det at spørsmålet kom med i Generalprøven, styrker min oppfatning.

## 7.3 Oppgaven Barnefødsler

I følge L97 ligger det i kravet til elevenes kunnskaper og ferdigheter at de skal tolke informasjoner presentert i tabeller og diagrammer.

> *Å behandle data er vesenlig med tanke på å skaffe oversikt over tallmateriale, å fremstille dette på en klar og riktig måte, både i tabeller og diagrammer, og trekke fornuftige konklusjoner fra dataene. I samfunnet gjøres det stadig mer bruk av grafiske framstillinger. Derfor må elevene lære å lese av og vurdere de informasjoner som et diagram eller en graf gir. (L97, s.157)*

Dette er ferdigheter de blir prøvet i til eksamen. Avgangsprøven for grunnskolen i både 2001 og 2003 presenterte en tidtabell i informasjonsheftet der elevene skulle bruke de gitte opplysninger for en videre reiseplan.

I rammeverket for PISA er *Kvantitativt resonnement* en av de fire sentrale ideer. Dette inkluderer å hente ut relevante tallinformasjoner. I dagliglivet er det oftest gjennom media elevene blir konfrontert med slike informasjoner, slik det også er i oppgaven *Barnefødsler*. Derfor vil konteksten til oppgaven plasseres i *Nærmiljø og samfunn*.

I oppgaven *Barnefødsler* som er utviklet rundt et oppslag i Dagsavisen, må elevene bearbeide de informasjoner som er gitt i et trediagram. De to første av de fire spørsmålene der elevene gjør bruk av *Kvantitative resonnement*, er det naturlig å plassere i kompetanseklasse 1, mens spørsmål 13 er fra kompetanseklasse 2.

I det siste spørsmålet vil elevene gjennom de informasjoner som er gitt i spørsmålet, møte usikkerhet som PISA-prosjektet for 2003 sterkt vektlegger. Dette spørsmålet er knyttet til statistikk og sannsynlighet i L97:

> *I opplæringen skal elevene:*
>
> - *arbeide med å utvikle mer presise begreper og uttrykksmåter for sannsynlighet og med å tallfeste sannsynligheter.(L97, s.169)*
>
> - *undersøke situasjoner der det må regnes med usikkerhet, risiko og sjanse. (ibid)*

Dette er et emne som faller vanskelig for mange elever. Problemtypene vil ofte være sammensatte og krever derfor originale løsningsstrategier. Det er derfor naturlig å sette spørsmål 14 i kompetanseklasse 3. En revidert utgave av oppgaven *Barnefødsler* der dette spørsmålet ble sterkt forandret, kom med i Generalprøven.

# BARNEFØDSLER

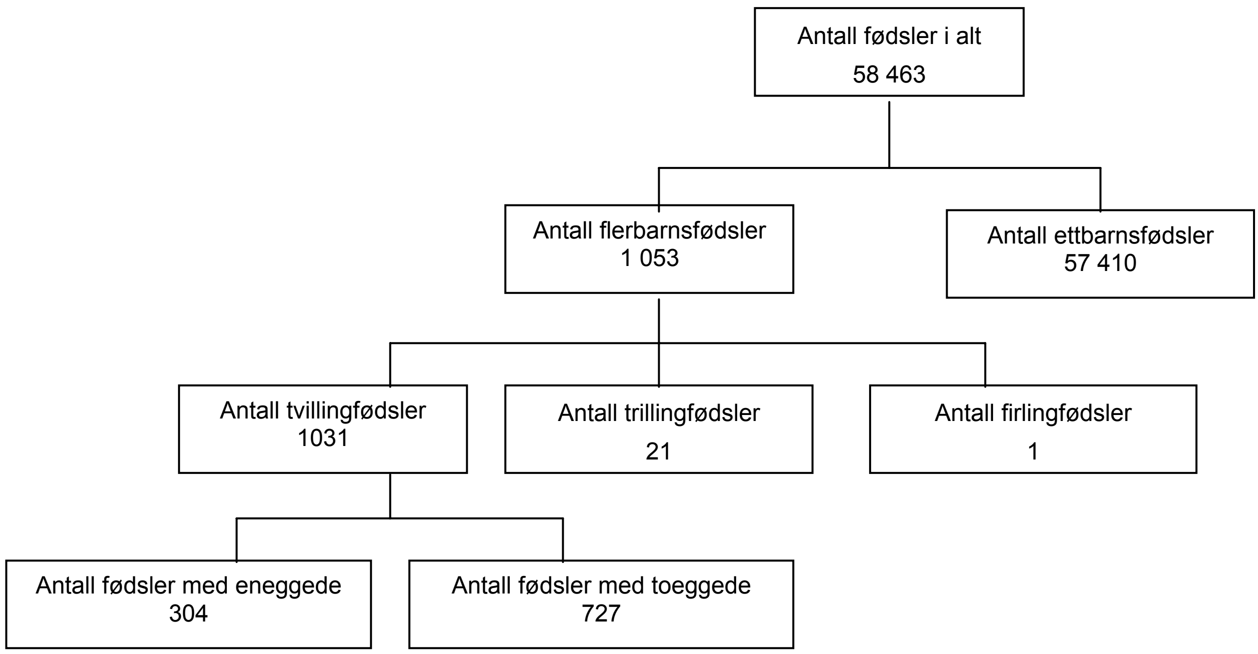
I Dagsavisen 2. oktober 2000 kunne vi lese:

## Tvillingfødsler doblet på 13 år

**Aldri før har så mange norske barn kommet til verden i par. Og aldri har så mange gravide hatt så store mager! På 13 år er antall tvillingfødsler doblet.**

Stadig flere barn deler plass i mors mage. I fjor ble det født dobbelt så mange tvillinger som i 1986. 18 av 1.000 fødsler var flerfødsler. Til sammen ble det født 1.031 tvillingpar i Norge. 21 trillinger så dagens lys og en mor bar til og med fram firlinger.

Antall barn født i flerbarnfødsler i 1999 i Norge var til sammen 2 129. Trediagrammet viser hvordan fødselsfordelingen var mellom ettbarnsfødsler og flerbarnsfødsler (tvillinger, trillinger osv):



*Les avisutklippet og diagrammet på foregående side og svar på spørsmålene som følger*

---

**Spørsmål 11:** BARNEFØDSLER

Hvor mange trillinger ble det født?

……………………….

---

For å kunne svare riktig på denne oppgaven bør elevene kunne:

- ha opparbeider ferdigheter til å kunne lese, formulere og formidle emner hvor det er naturlig å bruke matematikkens språk og symboler.
- finne og trekke ut informasjon fra tabell, diagram og annet datamateriale og drøfte eventuelle usikkerheter, skjevheter eller feilkilder.

I L97 heter det også

- *trene seg i å velge og vurdere framgangsmåter for å løse problemer i å gjøre overslag, kontrollere beregninger og vurder svar.(L97, s.168)*

# Koder til spørsmål 11

| **Koder** | **Svaralternativer** | **Frekvens** | **Prosent** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 63 | 57 | 39,9 | 0,554 | 0,45 |
| | **Ikke riktig svar** | | | | |
| 70 | 42 | 1 | 0,7 | 0,517 | 0,04 |
| 71 | 21 antall fødsler | 81 | 56,6 | -0,329 | -0,38 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 4 | 2,8 | -1,376 | -0,23 |

Slik spørsmålet var gitt, er det bare én kode for riktig svar på dette spørsmålet. Tabellen viser at p-verdien er 0,4. Dessuten diskriminerer den godt, "point-biserial" for riktig svar er 0,45.

Matematikken spørsmålet forventer at eleven skal beherske, er triviell. Man skulle derfor ha forventet en betydelig høyere p-verdi enn besvarelsene faktisk gir. Det må derfor finnes årsaker til den relativt lave p-verdien.

Kode 70 har også en høy positiv z-skåre, men "point-biserial" viser en verdi like over null. Den ene eleven som hadde dette svaralternativet multipliserte antall fødsler med to. Denne eleven svarte riktig på bare spørsmål 13 i denne oppgaven. Det må legges til at en z-skåre som er regnet ut på grunnlag av få elever, i dette tilfelle én, sier lite. Her ser vi nytten av den utregnede "point-biserial" korrelasjon for hvert svaralternativ. Dette er nærmere forklart ved innledningen til analysen i dette kapittelet.

Spørsmålet forutsatte at elevene leste ut antall trillingfødsler i trediagrammet for så å finne svaret ved å multiplisere antall fødsler med 3. Derved ville de få det antall trillinger som ble født det året, i utgangspunktet et ganske enkelt spørsmål. Det var antagelig mange av elevene som misforsto diagrammet eller spørsmålet ved at antall trillingfødsler i diagrammet ble oppfattet som antall trillinger. Derfor fikk nok kode 71 så stor oppslutning. Problemet lå kanskje i antallet 21 siden 7 trillingfødsler multiplisert med 3 barn i hver fødsel gir nettopp 21 trillinger. For å unngå et slikt problem ville en annen formulering av spørsmålet vært ønskelig. Et eksempel på det kan være: *Vi finner av diagrammet antall trillingfødsler. Hvor mange trillinger ble det født*? Det er sannsynlig at flere ville mestre spørsmålet dersom det ble spurt etter antall firlinger som ble født, siden det var én firlingfødsel det året. En revidert utgave av dette spørsmålet ble senere presentert for en gruppe elever der oppgaven nettopp var å finne antall firlinger som ble født, viste at p-verdien ble betraktelig høyere.

Av de 81 besvarelsene under kode 71, var det 25 med positiv z-skåre.

En vesentlig hensikt med den test PISA legger opp til er å fastslå i hvilken grad elevenes ferdigheter også er anvendelig utenfor en undervisningssituasjon. Spørsmålet som er relatert til en kontekst, i dette tilfelle et avisoppslag, inneholdt feller som dessverre altfor mange elever gikk i.

---

**Spørsmål 12:** BARNEFØDSLER

Hvor mange barn ble det totalt født det året?

…………………………

---

## Koder til spørsmål 12

| Koder | Svaralternativer | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 59539 | 40 | 28 | 0,745 | 0,47 |
| | **Ikke riktig svar** | | | | |
| 70 | 58463 antall fødsler | 87 | 60,8 | -0,209 | -0,26 |
| 79 | Andre svar | 10 | 7,0 | -0,406 | -0,11 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 6 | 4,2 | -1,257 | -0,26 |

Her gir koden for riktig svar en p-verdi på 0,28 som er noe lavere enn for foregående spørsmål. 72 % får null poeng enten ved at de svare feil eller lar være å svare. Vi ser igjen at 61 % leser ut antall fødsler direkte av diagrammet uten å ta hensyn til at antall fødsler selvfølgelig må gi et høyere antall barn som er født. Av de 87 besvarelsene under kode 70 var det så mange som 33 med positiv z-skåre.

Kode 10 har høy z-skåre på 0,745, som ligger betydelig over gjennomsnittelig totalskåre, som selvfølgelig alltid er null. Høy "point-biserial" viser at spørsmålet diskriminerer bra. Både **Ikke riktig svar** og **Ikke svart** har negativ z-skåre, slik det også bør være. Dette spørsmålet viste igjen i likhet med foregående spørsmål at mange elever ikke leste og behandlet informasjonen fra diagrammet riktig. Til Generalprøven fikk da også dette spørsmålet en innledning for å avverge den misforståelsen som mange ble offer for under pilottesten (se oppsummering p. 7.3.1).

Spørsmålene 11 og 12 er nært beslektet. En krysstabell viser om det var de samme personene som feilet på begge spørsmålene. Her er bare de som svarte på begge spørsmålene tatt med.

**Krysstabell mellom spørsmål 11 og spørsmål 12**

Count

| | | Spørsmål 12 | | | |
|---|---|---|---|---|---|
| | | 10 | 70 | 79 | Total |
| Spørsmål 11 | 10 | 38 | 14 | 4 | 56 |
| | 70 | 0 | 0 | 1 | 1 |
| | 71 | 2 | 73 | 5 | 80 |
| Total | | 40 | 87 | 10 | 137 |

Her kommer det fram at to av elevene som ikke svarte riktig på spørsmål 11, svarte riktig på spørsmål 12. Det er 79 elever som svarer feil på begge spørsmålene. Det vil si at 97,7 % av de elevene som ikke svarte riktig på spørsmål 11 heller ikke svarte riktig på spørsmål 12. Av disse var det 26 elever med positiv z-skåre på testen. Slår man sammen kode 70 og kode 79 på begge spørsmålene, gir det en krysstabell med en Kji-kvadratverdi på hele 67,78 som understreker en sterk sammenheng mellom svarmønsteret på disse to spørsmålene.

Problemet for mange var at de ikke hentet ut riktig informasjon. En frekvenstabell viser at av de 38 elevene som svarte riktig på begge spørsmålene, var det 11 med negativ z-skåre. Dette viser klart at mange "svake" elever ikke hadde problemer med å lese ut riktige informasjoner fra diagrammet.

Man kan da spørre seg om det var nødvendig å presentere to forholdsvis like spørsmål på samme oppgave, eller om det ene spørsmålet burde vært erstattet med en flervalgsoppgave. Det må her påpekes at i pilottesten er intensjonen å teste spørsmålene og ikke elevprestasjonen. Elevbesvarelsene er ett av grunnlagene for å teste hvorvidt de enkelte spørsmålene egner seg i en større testsammenheng. Med forholdsvis like spørsmål der den ene er gitt i åpent format og den andre i flervalgsformat, vil analysen av resultatet bidra til å fastslå hvilke av de to oppgaveformatene som best vil egne seg for senere bruk.

---

**Spørsmål 13:** BARNEFØDSLER

Hvor stor prosentandel av de nyfødte var tvillinger? Vis utregningene dine.

---

## Koder til spørsmål 13

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 20 | 3,46 % med utregning | 19 | 13,3 | 1,264 | 0,50 |
| 10 | 1,76 der oppsettet for prosentregning var riktig | 43 | 30,1 | 0,509 | 0,34 |
| | **Ikke riktig svar** | | | | |
| 70 | Kommafeil | 4 | 2,8 | -0,015 | 0,00 |
| 79 | Andre svar | 23 | 16,1 | -0,496 | -0,22 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 54 | 37,8 | -0,637 | -0,50 |

I dette spørsmålet skal elevene hente ut informasjon fra diagrammet for så å regne ut hvor stor prosentandel av de nyfødte som var tvillinger. Av de 62 elevene som viste riktig prosentregning, var det bare 19 som svarte helt riktig på oppgaven. Dette tilsvarer 13,3 % av de elevene som var med på testen. Dette lave prosenttallet grunner seg blant annet i at en forutsetning for å få helt riktig besvarelse var at de også taklet de to foregående spørsmålene. Derfor ble det benyttet koder for både helt riktig svar og delvis riktig for elever som viste riktig oppsett på prosentregning. Alle de som fikk kode 10, hadde regnet ut prosentandelen tvillingfødsler av samtlige fødsler. Her har vi igjen et eksempel på at mange elever ikke hentet ut de riktige informasjonene fra diagrammet.

Jeg har valgt å kode 70 for de 4 elevene som riktignok kunne sette opp riktig prosentregning, men kom ut med åpenbart feil prosent ved å sette komma på feil plass. Under kode 79, andre svar, ble de elevene plassert som helt tydelig ikke mestret denne form for prosentregning, å finne prosenten.

Som tidligere nevnt er en viktig side ved PISA at elevene blir prøvd i spørsmål der de skal hente ut relevante opplysninger fra eller tolke tabeller eller diagrammer. Siden dette samtidig blir sterkt presisert i L97, er det noe overraskende at så mange som 38 % ikke svarte på spørsmålet. På de to foregående spørsmålene var det svært få som ikke oppga et svar, så problemet for mange kan ligge enten i å finne prosenten eller en kombinasjon av å finne prosentandelen og det å hente ut relevante informasjoner fra diagrammet. De kan også ha vært så usikre på om de har svart riktig på foregående spørsmål at dette har medvirket til at de ikke oppga svar på de to siste spørsmålene i oppgaven.

Spørsmålet oppnådde en p-verdi på bare 0,28.

Tar man utgangspunkt i de som svarte helt riktig (kode 20) og delvis riktig (kode 10) på spørsmålet, er "point-biserial" på 0,69, som indikerer svært god diskriminering. Disse elevene fikk også en høy gjennomsnittlig z-skåre på henholdsvis 1,26 og 0,509. Av de 43 elevene som viste riktig prosentregning, men tolket diagrammet feil, var det 32 som hadde positiv z-skåre på testen.

En interessant side ved dette spørsmålet var å se på eventuelle forskjeller i hvordan jenter og gutter mestret den. I den følgende krysstabellen sammenlignes jentene og guttenes besvarelse på spørsmål 13.

**Krysstabell mellom kjønn og spørsmål 13**

Count

| | | Spørsmål 13 | | | | |
|---|---|---|---|---|---|---|
| | | 10 | 20 | 70 | 79 | Total |
| kjønn | Jenter | 25 | 16 | 3 | 7 | 51 |
| | Gutter | 18 | 3 | 1 | 16 | 38 |
| Total | | 43 | 19 | 4 | 23 | 89 |

Her fremkommer det en klar sammenheng mellom kjønn og besvarelse med bare 3 gutter mot 16 jenter for de som er kodet 20. Det gjelder de elevene som både har lest de relevante informasjonene riktig ut av diagrammet og har regnet ut riktig prosentandel. For de av elevene som bare hadde riktig oppsett på prosentregningen, var forskjellen ikke så stor. Ser vi på de elevene som oppnådde poeng på spørsmålet, var det 51,8 % av alle jentene og

33,8 % av alle guttene. Selv om jentene skåret ubetydelig bedre på hele testen, var den gjennomsnittlige z-skåren for de jentene som oppnådde poeng på spørsmålet noe lavere enn tilsvarende for guttene, med henholdsvis 0,661 og 0,864.

En Kji-kvadrattest på krysstabellen gir da følgende resultat. Her er tatt med 2 poeng for kode 20 og 1 poeng for kode 10. Ikke riktig gir 0 poeng.

**Kji-kvadrattest**

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 10,167[a] | 2 | ,006 |
| Likelihood Ratio | 10,840 | 2 | ,004 |
| Linear-by-Linear Association | 9,998 | 1 | ,002 |
| N of Valid Cases | 89 | | |

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 8,11.

Vi ser at den beregnede Kji-kvadrattesten gir en verdi på 10,17 med to frihetsgrader[15]. Sammenligner vi dette med den tilsvarende kritiske verdien på 5,99, kan vi konkludere med at det er en signifikant sammenheng mellom kjønn og hvordan spørsmål 13 ble besvart. Slår man sammen både kode 20 og kode 10 og kjører tilsvarende test, viser den også signifikante forskjeller.

Man kan ikke fastslå at jentene var bedre enn guttene til å lese ut de relevante informasjoner fra trediagrammet. En krysstabell på kjønn for både spørsmål 11 og spørsmål 12 viste da heller ingen signifikante forskjeller. Med andre ord, jentene mestrer prosentregningen på dette spørsmålet bedre enn guttene.

---

**Spørsmål 14:** BARNEFØDSLER

Tvillingpar kan enten være eneggede eller toeggede. 30 % av tvillingene er eneggede. De vil da alltid ha samme kjønn. For toeggede tvillinger er sannsynligheten for at disse er av samme kjønn lik 50 %. Totalt blant tvillingene, både eneggede og toeggede, er det like mange gutter som jenter.

På vei til skolen møter du ei jente som forteller deg at hun er tvilling.

Hvor stor prosent sannsynlighet er det for at hun har en tvillingsøster? Forklar hvordan du kom fram til dette svaret.

---

[15] Den kritiske verdien for Kji-kvadrattest avhenger også av størrelsen på tabellen den er utregnet på. I dette tilfelle har vi to kolonner for **Riktig svar** og ett for **Ikke riktig svar.** I tillegg har vi to rader for kjønn. Antall frihetsgrader blir $(3 - 1) \times (2 - 1) = 2$.

## Koder til spørsmål 14

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 20 | 65 % med forklaring | 13 | 9,1 | 1,127 | 0,36 |
| 10 | 65 % uten forklaring | 3 | 2,1 | 0,556 | 0,08 |
| | **Ikke riktig svar** | | | | |
| 70 | 50 % | 29 | 20,3 | 0,174 | 0,09 |
| 71 | 80 % | 15 | 10,5 | 0,123 | 0,04 |
| 72 | 40 % | 2 | 1,4 | 1,286 | 0,15 |
| 79 | Andre svar | 26 | 18,2 | -0,002 | 0,00 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 55 | 38,5 | -0,468 | -0,37 |

Her er det også brukt kode 20 for helt riktig svar med forklaring og kode 10 for riktig prosentandel, men der forklaringen er utelatt. De fleste som oppga riktig svar med forklaring, hadde summert 30 % (som er eneggede) og halvparten av 70 % (som er toeggede). De andre med riktige forklaringer gikk fram på samme måte, men her opererte man med andeler.

Man skulle anta at de 13 elevene under kode 20 representerte de som gjorde det best på hele testen med både høy ”point-biserial”, tatt i betraktning av antallet og den gjennomsnittlige

z-skåren. En frekvenstabell viser imidlertid at tre av disse elevene hadde negativ z-skåre. De tre besvarelsene under kode 10 hadde også relativt høy z-skåre, men en liten positiv ”point-biserial”. Dette viser, som det ble kommentert under spørsmål 11, at z-skåren på grunnlag av få elever ”sier lite” (se innledningen til dette kapittelet). Dette gjelder også for de to under kode 72 som hadde den høyeste gjennomsnittlige z-skåren. ”Point–biserial” utregnet for kode 20 og kode 10 under ett blir på 0,37. Dette er høyere enn for hver av kodene separat. Dette viser også p-verdiens betydning ved utregning av ”point-biserial”.

Under kode 70 var det hele 29 elever som feilaktig svarte 50 %, og enkelte av disse forklarte at det ble 50 % uavhengig av om de var en- eller toeggede. I utregningen som lå til grunn for besvarelsene under kode 71, tok de utgangspunkt i at alle de eneggede, som er 30 %, og halvparten av de som er toeggede, er jenter. Feilen var at de ikke tok halvparten av de 70 %, men halvparten av 100 % slik at de endte opp med 30 % og 50 % som til sammen blir 80 %. De to under kode 72 som svarte 40 % kom fram til denne prosenten ved at de summerte 50 % med 30 % og deretter dividerte på 2. For de besvarelsene som er kodet 79 er det vanskelig å finne et klart mønster for svaret de oppga. 55 elever svarte ikke på spørsmålet. Av disse var det igjen 33 som heller ikke svarte på det foregående spørsmål. Dette var gjennomgående svake elever der en frekvenstabell viser at 31 av disse hadde negativ z-skåre på hele testen.

Dette spørsmålet var det da også naturlig å plassere i kompetanseklasse 3 som krever matematisering av sammensatte problemstillinger og utvikling av originale løsningsmetoder. Spørsmålet fikk en p-verdi på 0,10, hvilket vil si at svært få elever mestret spørsmålet. De positive z-skårene på både kode 70, 71 og 72 viser at det var en stor grad av usikkerhet omkring hvordan dette spørsmålet skulle besvares. Det viser også at ved bruk i en senere testsammenheng bør dette spørsmålet først bearbeides.

### 7.3.1 Oppsummering rundt oppgaven Barnefødsler

Det var bare fem elever som svarte riktig på alle fire spørsmålene. Med gjennomsnittlig z-skåre og ”point-biserial” på henholdsvis 1,676 og 0,32 mener jeg at også denne oppgaven i sin helhet, i likhet med oppgaven *Trappekonstruksjon*, diskriminerte godt mellom de aller flinkeste og de andre.

Det var fire elever som ikke svarte på noen av spørsmålene i denne oppgaven. Her ser vi igjen at oppgaver med kontekst ”nær” elevene øker svarprosenten.

Et av målene i PISA er å finne ut hvordan 15-åringer behersker tekster de møter i hverdagen. Dette betyr også å kunne trekke ut og bruke de opplysninger som ligger i teksten enten det handler om å tolke tekster eller å hente ut relevante opplysninger. Derfor har PISA utviklet et rammeverk for reading literacy med følgende definisjon:

> *The capacity to understand, use, and reflect on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and participate in society. (OECD 2000 s.10)*

Et viktig aspekt ved det å mestre en tekst er to retrieve information som kan oversettes med å hente ut relevante informasjoner fra en tekst. Dette er en side ved lesekompetansen som riktignok er et annet område for PISA-undersøkelser enn det jeg har vektlagt. Funksjonelle ferdigheter som det å lese ut av et diagram, har like stor gyldighet også i matematikkdelen i PISA. En utvidet definisjon av numeracy utover betydningen av de grunnleggende regneferdigheter, er gitt i Cockcroft-rapporten. Den inkluderer også betydningen av å mestre informasjoner presentert i for eksempel et diagram.

> *The first of these is an ”at-homeness” with numbers and an ability to make use of mathematical skills which enables an individual to cope with the practical mathematical demands of his everyday life. The second is an ability to have some appreciation and understanding of information which is presented in mathematical terms, for instance graph, charts or tables or by reference to percentage increase or decrease. (Cockcroft 1982, s.11)*

Til Generalprøven ble oppgaven *Barnefødsler* noe omarbeidet selv om konteksten ble den samme. Spørsmål 12 fikk en innledende tekst som presiserte at diagrammet oppga antall fødsler. Der sto det følgende:

> *Den øverste ruta i diagrammet viser at det var 58463* ***fødsler*** *totalt i 1999. Dette er imidlertid* ***ikke*** *det totale antallet barn som ble født i 1999, fordi i hver flerbarnsfødsel blir det født mer enn ett barn.*
>
> *Hvor mange barn ble det totalt født i Norge i 1999?*

Med denne presiseringen burde spørsmålet få en betraktelig høyere p-verdi. Til Generalprøven ble både spørsmål 13 og 14 erstattet med et nytt spørsmål der konteksten til spørsmål 14 gjennomgikk en forandring. Spørsmålet som i utgangspunktet hadde for høy vanskegrad, ble erstattet med tre noe enklere spørsmål med henvisning til diagrammet. Her skulle elevene regne ut både prosentandelen jentetvillinger, guttetvillinger og jente-gutt-tvillinger av totalt antall tvillinger. Spørsmålene er fortsatt plassert under Usikkerhet i rammeverket for PISA.

Ved oppgaven *Barnefødsler* ble elevene presentert et diagram fra media som de i dagliglivet ofte blir presentert med. Problemstillingen knytter seg nær opp til constructet ved *mathematical literacy* i PISA og L97 slik det er sitert i p.1.2.1.

## 7.4 Oppgaven Fahrenheit – Celsius

For å opprettholde elevenes motivasjon og interesse var det igjen på tide med spørsmål med forventet høy svarprosent. Selv om konteksten til oppgaven *Fahrenheit – Celsius* er lokalisert i distanse lengst fra elevene, *Vitenskapelig*, regnet jeg med at den er enkel å oppfatte[16]. Den innledende teksten er kort og problemstillingen er fra kompetanseklasse 1 for de to første spørsmålene og 2 for de fire neste.

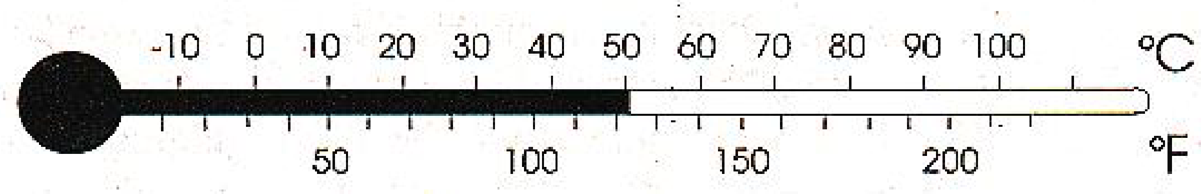
For å løse spørsmålene i denne oppgaven må elevene se sammenhengen mellom fahrenheitskalaen og celsiusskalaen enten gitt ved et termometer som viser begge skalaene, eller ved hjelp av en oppgitt omregningsformel mellom disse to måleenhetene. Det er naturlig å plassere oppgaven med unntak av spørsmål 15 og 16, i det som rammeverket for PISA har organisert i *Forandring og sammenheng*. Spørsmål 15 og 16 vil jeg ut fra den matematikken som ligger i spørsmålet, plassere i området *Kvantitativt resonnement*.

[16] Både PISA og *Australian Council for Education* (ACER) har i sin pilottest plassert oppgaven under *Vitenskapelig* kontekst. Jeg anser selv at den for mange elever ligger i grenseområdet til *Nærmiljø og samfunn*, altså situasjon ”nærmere” eleven.

# FAHRENHEIT – CELSIUS

Det er i hovedsak to typer skalaer som brukes i dag for å måle temperatur, fahrenheitskalaen og celsiusskalaen. På fahrenheitskalaen blir vannets frysepunkt satt til $32^0$ og normal kroppstemperatur hos et friskt menneske er $96^0$. På celsiusskalaen er vannets frysepunkt satt til $0^0$ og vannets kokepunkt er satt til $100^0$.

I de fleste land er det celsiusskalaen som brukes, mens fahrenheitskalaen brukes først og fremst i USA. I noen få land hvor celsiusskalaen er den offisielt vedtatte, blir likevel fahrenheitskalaen ofte brukt.




Figuren ovenfor viser et termometer med både celsiusskalaen ($^0$C) og fahrenheitskalaen ($^0$F).

*Svar på følgende spørsmål ved å lese av på figuren:*

**Spørsmål 15:** FAHRENHEIT – CELSIUS

En genser har et merke hvor det står at den skal vaskes på 50 $^0$C. Hva blir det i $^0$F?

………$^0$F

Kravet til å godkjenne svaret som riktig, ble satt til at den angitte temperaturen ligger mellom 121 og 124 fahrenheitgrader.

## Koder til spørsmål 15

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | $121^0$F – $124^0$F | 99 | 69,2 | 0,282 | 0,42 |
| | **Ikke riktig svar** | | | | |
| 70 | $125^0$F – $127^0$F | 6 | 4,2 | -1,159 | -0,24 |
| 71 | $118^0$F – $120^0$F | 19 | 13,3 | -0,467 | -0,18 |
| 79 | Andre svar | 13 | 9,1 | -0,320 | -0,10 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 6 | 4,2 | -1,317 | -0,28 |

Svarprosenten på dette spørsmålet ble hele 95,8 % som viser at intensjonen om at mange ville prøve seg på denne, ble innfrid. Selv om koden for riktig svar fikk en p-verdi på 0,69 diskriminerte den godt med "point-biserial" på 0,42, og en gjennomsnittlige z-skåren på bare 0,282. Med noe større toleranse for riktig svar (flere av besvarelsene som ble kodet under 70 og 71) ville p-verdien vært høyere. Alle besvarelser bortsett fra kode 10 fikk negativ gjennomsnittlig z-skåre. Selv om de 6 elevene under kode 70 var blant testens svakeste, var deres avlesing av temperaturskalaen bare noen få grader over det riktige svaret.

Under kode 79 finner vi blant annet de som vegret seg for "den høye vasketemperaturen" selv om den da var i Fahrenheit. Her var også de som gikk den andre veien, fra Fahrenheit til Celsius, og oppga $10^0$C. Det var også to besvarelser som bommet på punktoppdelingen av fahrenheitskalaen mellom 100 og 150. Disse oppga et svar litt over 100.

Det kan vurderes om jeg burde ha delt opp kode 79 slik at de to som bommet på punktoppdelingen fikk egen kode. Noen elever har problemer med å fastsette punkter på en skala. Om testen hadde som oppgave å identifisere slike problemer, ville det vært naturlig å ha en egen kode for slike. I utgangspunktet vurderte jeg en slik egen kode, men siden så få besvarelser kom under denne ble de lagt inn i kode 79. Dette støttes av at disse oppgaveforslagene ikke er forslag til en diagnostisk test.

---

**Spørsmål 16:** FAHRENHEIT – CELSIUS

Temperaturen en dag er 100 $^0$F. Hva blir det i $^0$C?

………$^0$C

---

Med en tilstrekkelig grad av toleranse ble alle forslag fra 36 $^0$C til 39 $^0$C godkjent som riktig svar.

## Koder til spørsmål 16

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 36 $^0$C - 39 $^0$C | 127 | 88,8 | 0,113 | 0,32 |
| | **Ikke riktig svar** | | | | |
| 70 | 40 $^0$C - 42 $^0$C | 4 | 2,8 | -0,696 | -0,12 |
| 71 | 33 $^0$C - 35 $^0$C | 6 | 4,2 | -0,745 | -0,16 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 6 | 4,2 | -1,179 | -0,25 |

Dette spørsmålet er svært likt spørsmål 15, men her skal elevene gå fra fahrenheitskalaen til celsiusskalaen. Svarprosenten ble den samme, men p-verdien ble hele 0,89. Med en så høy p-verdi ble også den gjennomsnittlige z-skåren og "point-biserial" for kode 10 lavere, med henholdsvis 0,113 og 0,32. Både "point-biserial" og den gjennomsnittlige z-skåren ble negativ for de andre svaralternativene. Selv de 10 elevene som ble kodet 70 og 71 forsto antagelig avlesningen, men var unøyaktige. En krysstabell viser at fem av de seks elevene under koden **Ikke svart**, heller ikke besvarte foregående spørsmål. Fire av disse var testens "svakeste" elever.

Det kan være flere årsaker til den store forskjellen på p-verdiene for to så like spørsmål. På spørsmål 15 måtte elevene forholde seg til en tekst som for enkelte antagelig skapte en viss usikkerhet rundt temperaturen. En annen årsak kan være skalaene på termometeret.

Betrakter man termometeret nærmere var det nok enklere for elevene å forholde seg til oppdelingen på celsiusskalaen der hver 10-ende grad er oppgitt. Det er interessant å observere at ingen besvarelser her ble kodet under andre svar.

---

En omregningsformel mellom disse to måleenhetene er

$$5x - 9y = 160,$$

der x er antall grader målt i Fahrenheit ($^0$F) og y er antall grader målt i Celsius ($^0$C)

**Spørsmål 17:** FAHRENHEIT – CELSIUS

Hvor mange grader får vi på fahrenheitskalaen når vi har 75 $^0$C?

………$^0$F

Avgangsprøven gitt til ungdomskoleeksamen få uker etter at pilottesten ble gjennomført, hadde også med en oppgave der elevene skulle omskrive omregningsformelen fra celsiusgrader til fahrenheitgrader.

For å gå i kast med dette spørsmålet:

- *bør elevene være kjent med bruk av bokstaver for å symbolisere variable tall og størrelser, og kunne se dette i sammenheng med formler og uttrykk. (L97, s.169)*

# Koder til spørsmål 17

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 167 $^0$F | 31 | 21,7 | 1,227 | 0,65 |
| | **Ikke riktig svar** | | | | |
| 70 | 168 $^0$F – 170 $^0$F | 66 | 46,2 | -0,202 | -0,19 |
| 71 | 164 $^0$F – 166 $^0$F | 16 | 11,2 | -0,533 | -0,19 |
| 79 | Andre svar | 17 | 11,9 | -0,402 | -0,15 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 13 | 9,1 | -0,721 | -0,23 |

Spørsmålet fikk en svarprosent på nesten 91. Siden omregningsformelen nå var oppgitt, var intensjonen at denne skulle brukes, og ikke at elevene nok en gang skulle lese av termometeret. Det var derfor bare det korrekte svaret som ble godkjent, selv om kanskje noen ”traff” riktig ved å lese av termometeret. Av de som svarte, var det 23,9 % som ble kodet under riktig svar. Disse hadde svært høy gjennomsnittlig z-skåre på 1,227 og ”point-biserial” på 0,65. Z-skåren er høy tatt i betraktning antallet. Det var bare én elev som hadde negativ z-skåre på hele testen. Dette viser at elevene som kom under denne koden, var blant de flinkeste. En frekvenstabell viser også at 19 av de 25 som skåret best på testen oppga riktig svar på dette spørsmålet. Selv om p-verdien ble lav, diskriminerte spørsmålet meget tilfredsstillende.

Siden bare svaret skulle oppgis og ikke fremgangsmåten, er det ikke sikkert at elevene benyttet omregningsformelen for å svare på dette spørsmålet. Det store antall besvarelser kodet 70 og 71 viser at flere benyttet termometeret for å gi et svar. Disse fikk en negativ gjennomsnittlig z-skåre, selv om de oppga en temperatur nær den riktige. Det betyr nødvendigvis ikke at de ikke mestret formelen, men like gjerne at de valgte ”den lette varianten”. Dersom oppgaven ble presentert som langsvaroppgave der det ble krevd enten utregning eller forklaring, eller om det sterkere ble presisert at omregningsformelen skulle benyttes, vil jeg anta at antall besvarelser under kode 70 og 71 ville blitt betraktelig redusert. Dette ville redusert svarprosenten, siden flere ville unnlat å svare på spørsmålet.

---

**Spørsmål 18:** FAHRENHEIT – CELSIUS

Hvor mange grader får vi på celsiusskalaen når vi har 5 $^{0}$F?

………$^{0}$C.

---

På forrige spørsmål tok muligens noen elever det ”lettvinte” avlesningsalternativet, mens de her i større grad ble tvunget til å benytte omregningsformelen siden det her var vanskeligere å lese ut temperaturen helt til venstre på termometeret.

# Koder til spørsmål 18

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig** | | | | |
| 10 | -15$^{o}$C | 56 | 39,2 | 0,589 | 0,47 |
| | **Ikke riktig svar** | | | | |
| 70 | -12$^{0}$C - -14$^{0}$C | 22 | 15,4 | -0,488 | -0,21 |
| 71 | -16$^{0}$C - -18$^{0}$C | 18 | 12,6 | -0,364 | -0,14 |
| 72 | 15$^{o}$C | 2 | 1,4 | 0,931 | 0,11 |
| 79 | Andre svar | 23 | 16,1 | -0,080 | -0,04 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 22 | 15,4 | -0,714 | -0,31 |

Intensjonen var her i likhet med foregående spørsmål at omregningsformelen skulle benyttes. Selv om spørsmålene var relativt like og elevene benyttet de samme operasjonelle kunnskaper, var svarprosenten her noe lavere. Til gjengjeld ble p-verdien betraktelig høyere, 0,39. Med et større antall riktige svar ble både gjennomsnittlig z-skåre på 0,589 og ”point-biserial” på 0,47 for kode 10, en del lavere enn for spørsmål 17. Den lavere gjennomsnittlige z-skåren skyldes i hovedsak at 19 av de 56 som oppga riktig svar, hadde negativ z-skåre på testen. I tillegg svarte 21 av de 25 flinkeste elevene riktig på spørsmålet.

Det var antagelig færre som leste av termometeret for å finne svaret på dette spørsmålet enn på det forrige, siden vi her beveget oss helt i termometerets yterkant. Det er likevel usikkert om alle besvarelser kodet 10 hadde benyttet omregningsformelen, selv om det er vanskelig å lese ut -15 $^{0}$C direkte på termometeret. En krysstabell viser overraskende nok at selv om 56 elever svarte riktig, var det bare 22 av disse som også svarte riktig på spørsmål 17. Dette var elever med en svært høy gjennomsnittlig z-skåre på hele 1,324. Dette tyder igjen på at flere her ble tvunget til å benytte omregningsformelen. Det kan argumenteres for at en krysstabell fungerer som en kontroll av hvorvidt elevene brukte eller mestret omregningsformelen fullt ut.

Det er naturlig å anta at de elevbesvarelser som ble kodet 70 og 71 i det vesentlig hadde benyttet termometeret. Antallet ble betydelig lavere enn for spørsmål 17.

Jeg valgte å plassere de to besvarelsene som oppga 15 $^{0}$C under egen kode. Krysstabellen for spørsmål 17 og 18 viser at de hadde rett svar på forrige spørsmål. De hadde også den høyest gjennomsnittlige z-skåren for alle svaralternativene, mens "point-biserial" var på bare 0,11. Her er det også igjen et eksempel på at z-skåre basert på få besvarelser sier lite (se kommentarer under spørsmål 11). Antagelig benyttet de omregningsformelen for begge svarene, men kom her ut med feil fortegn. De har heller ikke foretatt den meget enkle kontrollen av om svaret kunne være riktig ved å se på termometeret.

Under kode 79 var det mange besvarelser det var vanskelig å finne et klart mønster i, bortsett fra de få tilfellene der noen elever gikk fra celsiusskalaen til fahrenheitskalaen og dermed oppga 40 grader (41 grader er riktig ved bruk av omregningsformelen). Den gjennomsnittlige z-skåren for besvarelsene under kode 79 viste med –0,08 at det her var elever som skåret nær det totale gjennomsnittet, som alltid er null. Dette kan tyde på at mange oppga et svar uten å regne ut. Trolig ville flere elever ved et annet oppgaveformat eller presisering av at omregningsformelen skulle benyttes (se oppsummeringen p 7.4.1), håndtert spørsmålet annerledes.

De 13 elevene som ikke besvarte spørsmål 17, svarte heller ikke på dette spørsmålet. De fikk nå følge av ytterligere 9 elever. Det er en klar tendens til at svarprosenten på spørsmålene i oppgaven *Fahrenheit – Celsius*, er avtagende.

En Kji-kvadrattest basert på Krysstabellen for spørsmål 17 og 18 viser for de som svarte riktig på begge spørsmålene, de som svarte riktig på ett av spørsmålene, og de som ikke svarte riktig på noen av spørsmålene, en verdi på 11,74. Dette viser en klar sammenheng mellom besvarelsene på spørsmålene.

**Spørsmål 19:** FAHRENHEIT – CELSIUS

To naboer med forskjellige termometre møtes. Den ene har et termometer som viser $^0$C og den andre et som viser temperaturen i $^0$F. Den ene sier til den andre: "I dag er det fem ganger så varmt for meg som for deg"

Når vil fahrenheitskalaen vise "fem ganger så mange grader" som celsiusskalaen? Vis ved å sette en ring rundt det riktige svaret

A 15 $^0$C

B 70 $^0$F

C 50 $^0$F

D 20 $^0$C

# Koder til spørsmål 19

Her er riktig svar merket med *.

| **Koder** | **Frekvens** | **Prosent** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| A | 11 | 7,7 | -0,107 | -0,03 |
| B | 12 | 8,4 | -0,617 | -0,19 |
| C* | 87 | 60,8 | 0,335 | 0,42 |
| D | 5 | 3,5 | -0,264 | -0,05 |
| **Ikke svart** | | | | |
| 99 | 28 | 19,6 | -0,674 | -0,34 |

Denne flervalgsoppgaven fikk en svarprosent på over 80. P-verdi på 0,61 viste at de fleste ikke hadde problemer med å velge riktig alternativ. De tre distraktorene fikk en samlet tilslutning av bare 28 elever, eller under 25 % av de som besvarte spørsmålet. Det tyder på at for mange elever var det lett å eliminere, og da spesielt distraktor D som fikk tilslutning av under 5 %. Alle distraktorene fikk negativ gjennomsnittlig z-skåre som selvfølgelig er ønskelig når p-verdien ble så høy.

Tar man i betraktning den høye p-verdien her i forhold til spørsmål 17, har antagelig mange elever som valgte riktig alternativ benyttet termometeret. Der står 10 celsiusgrader direkte overfor 50 fahrenheitgrader. Her burde minst én eller flere av distraktorene erstattes med nye som oppga en temperatur nær 10 celsiusgrader eller 50 fahrenheitgrader.

En annen mulighet er å erstatte oppgaven med en annen flervalgsoppgave hvor det ikke er like lett å lese ut svaret av termometeret. Forslag til nye spørsmål kan være:

- *Når vil fahrenheitskalaen vise dobbelt så mye som celsiusskalaen?*
- *I en badstue er det et termometer som viser celsiusgrader og et som viser fahrenheitgrader. Hvilken temperatur er det i badstua når fahrenheittermometeret viser 100 grader mer enn celsiustermometeret?*

Disse to spørsmålene vil gjøre det vanskeligere å lese svaret ut av termometeret som er presentert i innledningen, og tvinger dermed elevene til å forholde seg til omregningsformelen. Dette vil gi større utfordringer og sannsynligvis senke p-verdien på spørsmålet betraktelig.

---

**Spørsmål 20:** FAHRENHEIT – CELSIUS

På Antarktis er det en internasjonal forskningsstasjon. En dag viste et fahrenheittermometer og et celsiustermometer "like mange" grader.

Hva var da temperaturen?

……………

---

## Koder til spørsmål 20

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | -40 | 5 | 3,5 | 1,463 | 0,28 |
| | **Ikke riktig svar** | | | | |
| 70 | -50 - -41 | 6 | 4,2 | 0,616 | 0,13 |
| 71 | -39 - -30 | 6 | 4,2 | 0,261 | 0,06 |
| 72 | $0^0$C | 10 | 7 | -0,524 | -0,14 |
| 73 | $10^o$C og $50^0$F | 6 | 4,2 | 0,064 | 0,01 |
| 74 | "Det vil ikke gå" | 3 | 2,1 | 0,478 | 0,07 |
| 79 | Andre svar | 24 | 16,8 | -0,089 | -0,04 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 83 | 58 | -0,084 | -0,10 |

Av en svarprosent på 42, var det bare 8,3 % av de som besvarte spørsmålet som ble kodet under **Riktig svar**. Spørsmålets p-verdi ble bare 0,04, som igjen betyr at spørsmålet falt vanskelig for elever i denne årsklassen. Gjennomsnittlig z-skåre viser at det var de flinkeste elevene som greide oppgaven. I motsetning til spørsmål 19, var det her ikke mulig å lese svaret på termometeret, så elevene var tvunget til å benytte omregningsformelen. Få elever så muligheten som lå i formelen. Settes y = x, fremkommer en enkel ligning:

$$5x - 9x = 160$$

$$-4x = 160$$

$$x = -40$$

Muligens var oppmuntringen flere oppfattet ved ”snarveien” de nok brukte for å svare på spørsmål 19, årsaken til at relativt mange forsøkte seg på oppgaven, men også feilet. Derfor valgte jeg å opprette flere koder for de som ikke svarte riktig. De 12 elevbesvarelsene som ble kodet 70 og 71 ga et svar opp til det riktige. Det kan spekuleres over hvordan de kom fram til svaret, enten ved å måle seg fram eller gjette en temperatur nær det riktige svaret. De har en gjennomsnittlig z-skåre på 0,439, som er betydelig over gjennomsnittet.

Kode 72 representerer de 10 besvarelsene som oppga $0^0$. Man kunne muligens anta at disse elevene, siden de hadde en sterk negativ gjennomsnittlig z-skåre, brukte termometeret og hadde problemer med oppmerkingen på fahrenheitskalaen. Det er også en mulighet for at de antok at begge skalaene viste likt ved $0^0$. Dette kan likevel ikke være årsaken for alle under denne koden. En krysstabell mellom spørsmålene 18 og 20 viser at 5 av de 10 elevene mestret punktoppdelingen på spørsmål 18. Det er mer sannsynlig at de etter å ha prøvd seg på spørsmålet uten å lykkes, valgte å gi et svar framfor å svare blankt.

De 6 elevbesvarelsene under kode 73 som oppga svaret ”10 celsiusgrader - 50 fahrenheitgrader”, har en liten positiv gjennomsnittlig z-skåre. Det er vanskelig å si hvordan de kom til dette svaret. De kan enten ha misforstått spørsmålet ved heller å vise når disse temperaturene viser likt med oppgitte grader, eller de kan ha overført svaret fra foregående oppgave. En krysstabell mellom spørsmålene 19 og 20 viser at bare 4 av disse svarte riktig på spørsmål 19, så denne muligheten kan ikke gjelde for alle.

Kode 74 var for de 3 elevene som mente at det var umulig for fahrenheittermometeret og celsiustermometeret å vise ”like mange” grader. Dette er elever med relativt høy gjennomsnittlig z-skåre. Trolig har de forholdt seg til det begrensede intervall som termometerfiguren i innledningen viste.

Spørsmålet ville tjent på om det var blitt bedt om forklaring eller begrunnelse for det svar som ble gitt. Kanskje ville svarprosenten blitt noe lavere, men gevinsten ville være at svarene kunne gitt informasjon om elevene hadde forstått oppgaven og hvordan de tenkte.

Slik spørsmålet her var gitt, er det lite egnet i en større testsammenheng, både på grunn av den lave p-verdien og den usikkerhet den antagelig skapte hos noen elever. Dette gjelder ikke minst for de elevene som ensidig forholdt seg til termometerskalaene.

### *7.4.1 Oppsummering rundt oppgaven Fahrenheit – Celsius*

Denne oppgaven besto av 6 spørsmål der noen var forholdsvis like. På de fem første spørsmålene var svarprosenten høy. Både spørsmålene 15 og 16 hadde høy p-verdi. Av de 52 elevene som svarte på alle spørsmålene i oppgaven, var det bare to elever som svarte riktig på alle. De hadde gjennomsnittlig z-skåre på hele 1,641 for testen, og verdien for ”point-biserial” ble på bare 0,20 grunnet de få elevene som svarte riktig på alle spørsmålene. Det var spesielt spørsmål 20 som var årsaken til at så få elever svarte riktig på alle spørsmålene. Ser man bort fra spørsmål 20, var det 17 elever som besvarte de andre spørsmålene riktig. De hadde samtlige positiv z-skåre med gjennomsnitt på 1,394. Under testen fikk jeg inntrykk av at oppgaven engasjerte elevene. Det viste seg også at bare fem elever ikke besvarte noen av spørsmålene. Disse var også svake med gjennomsnittlig z-skåre på –1,376.

Intensjonen var at elevene skulle benytte termometerskalaen for å svare på de to første spørsmålene og omregningsformelen på de neste. Det viste seg at flere besvarelser ikke forholdt seg til omregningsformelen der det var meningen at den skulle benyttes.

I ettertid ser jeg at de fire siste spørsmålene kunne erstattes med en sammensatt flervalgsoppgave der elevene skal svare på en rekke påstander. Etter presentasjonen av omregningsformelen bør det presiseres at elevene skal benytte formelen for å besvare hver av de etterfølgende påstander. Eksempler på slike påstander kan være:

- *Når celsiusskalaen viser 0 grader, viser fahrenheitskalaen 31 grader.*
- *Det er én temperatur hvor begge skalaene viser samme gradetall.*
- *Det er én temperatur der celsiustermometeret viser "fem ganger så mange grader" som fahrenheittermometeret.*

Dersom disse spørsmålene også skal fungere som en diagnostisk oppgave (noe som ikke var intensjonen), bør elevene også begrunne sine valg. På den måten ville de gi informasjon både om i hvilken grad de har forstått påstandene, og framgangsmåten de har benyttet.

I artikkelen *Justifying the selection of answers in multiple choice items* argumenterer Tamir (1990, s.563-573) for dette. Hans argumentasjon gjelder presentasjon av flervalgsoppgaver med bruk av distraktorer, men er også relevant for sammensatte flervalgsoppgaver. I tillegg til de opplysningene som fremkommer ved at elevene begrunner valg av alternativene, enten det er de enkle eller de sammensatte flervalgsoppgavene, blir elevene kanskje i større grad tvunget til å overveie alternativene. På den måten blir de forhåpentligvis konfrontert med sine eventuelle misoppfatninger (se p. 3.2) under testen.

Denne oppgaven kom ikke med til Generalprøven. Argumenter mot å bruke oppgaven i en internasjonal test er at i en del land der celsiusskalaen er den offisielt vedtatte, blir likevel fahrenheitskalaen ofte brukt. På det vis ville elevene i disse land ha et fortrinn framfor de land som bruker bare den ene skalaen.

Jeg anser at denne oppgaven, noe omarbeidet, kan fremstå som en god testoppgave. Til avgangsprøven i grunnskolen 2001 noen uker etter at pilottesten ble gjennomført, var det med en oppgave der elevene skulle finne en formel for Fahrenheit uttrykt ved Celsiusgrader (Læringssenteret 2001b).

## 7.5 Oppgaven Fibonaccitallene

Denne oppgaven har i likhet med innledningen til oppgaven *Fahrenheit – Celsius* en vitenskaplig kontekst. Men i motsetning til *Fahrenheit – Celsius*, kan den oppfattes som mer abstrakt, selv om innledningen, som var mer omfattende, forsøkte å engasjere elevene ved hjelp av fenomener vi stadig finner i naturen. I vinklingen til hvordan fibonaccitallene fremkommer, er det tatt utgangspunkt i å tegne kvadrater. Derfor blir det naturlig å plassere spørsmål 21 under *Rom og form*. Siden elevene videre i oppgaven skal utvikle en tallfølge, plasseres de tre neste spørsmålene i *Forandring og sammenheng*.

Min oppfatning er at alle spørsmålene i oppgaven bør plasseres i kompetanseklasse 2 selv om *Australian Council for Education Research* (ACER) hadde foreslått spørsmål 23 i kompetanseklasse 3.

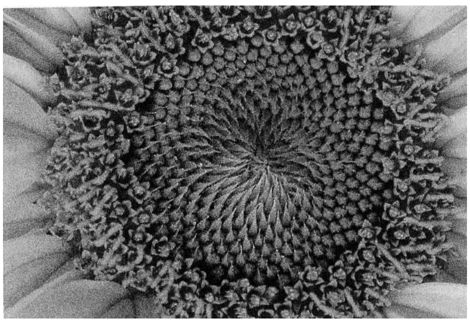
---

# FIBONACCITALLENE

Leonardo fra PISA levde i årene 1175 til 1225. Han var mer kjent under navnet Fibonacci.

Han er i dag kjent i matematikken fordi han fikk en tallfølge oppkalt etter seg. Denne tallfølgen, Fibonaccitallene, finner vi stadig igjen i naturen. Mange vekster danner eksempelvis spiraler. Ser vi på grankonglens dekkskjell er det lett å følge slike spiraler, og i hver retning er det flere. Teller vi antall spiraler i hver retning finner vi at disse alltid er enten 3 den ene veien og 5 den andre, eller 5 og 8, eller 8 og 13 osv. Dette er alle par av tall som følger rett etter hverandre på rekken av Fibonaccitall.

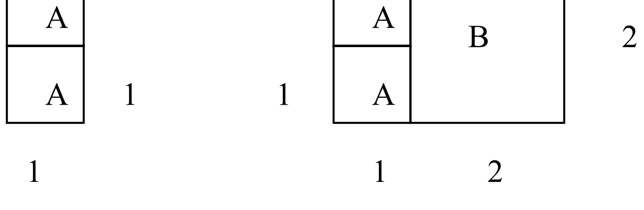
Tar vi for oss en solsikke (se bildet av solsikken nedenfor) danner frøene 21 spiraler hvis vi går mot urviseren, og 34 hvis vi går med urviseren.



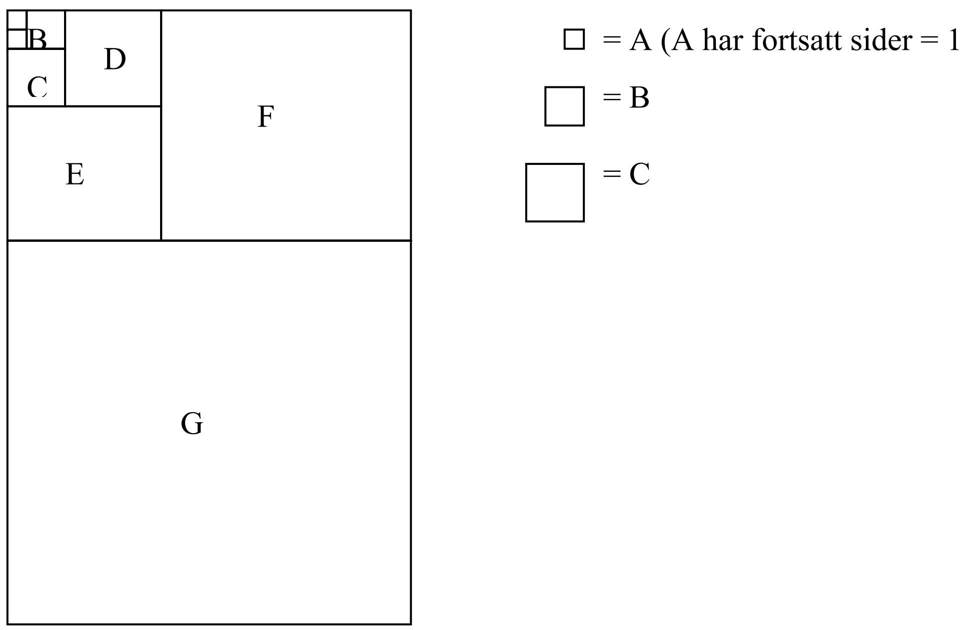
Det finnes også solsikker med for eksempel 55 og 89 spiraler, som også er to Fibonaccitall som følger rett etter hverandre.

Vi kan se hvordan Fibonaccitallene fremkommer ved å tegne kvadrater. Vi starter med et kvadrat som har sider 1. Opp til denne setter vi et nytt kvadrat, som også har sider 1.

Til disse setter vi et nytt kvadrat som har sider lik summen av de to foregående.



Vi setter av et nytt kvadrat med sidelengde som er summen av de to foregående, og slik fortsetter vi til vi får følgende figur:




**Spørsmål 21:** FIBONACCITALLENE

Fyll ut kvadratenes sider i tabellen nedenfor

| **Kvadratene** | A | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|---|
| **Sidene i kvadratene** | 1 | 1 | 2 | | | | | |

I følge L97 bør elevene:

- *kjenne til viktige egenskaper ved geometriske objekter. De skal kunne lage figurer, foreta avbildninger og skape mønstre.(L97, s.166)*

- *arbeide med spennende sammenhenger fra tallenes verden, f. eks. tall med spesielle egenskaper, den rolle tallmystikk kan spille i enkelte kulturer, eller den tiltrekning tallgåter kan ha. (ibid, s.170)*

## Koder til spørsmål 21

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 3,5,8,13,21 | 72 | 50,3 | 0,717 | 0,73 |
| | **Ikke riktig svar** | | | | |
| 70 | Ubetydelig feil | 3 | 2,1 | -0,429 | -0,06 |
| 71 | Potensutvikling | 5 | 3,6 | -0,500 | -0,10 |
| 72 | 3,4,5,6,7 | 9 | 6,3 | -0,942 | -0,25 |
| 73 | 22 33 44 | 3 | 2,1 | -0,508 | -0,08 |
| 79 | Andre svar | 18 | 12,6 | -0,574 | -0,22 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 33 | 23,1 | -0,834 | -0,46 |

Introduksjonen til oppgaven *Fibonaccitallene* var omfattende. Det var derfor en fare for at flere elever ville hoppe over den. Derfor anser jeg svarprosenten, 77, på det første spørsmålet i oppgaven som tilfredsstillende. Av de som svarte var det 70 % som kom under kode 10. De tok utgangspunkt i sidene i kvadratene som har lengder fra *Fibonaccitallene*. Spørsmålet fikk en p-verdi på 0,5 med en gjennomsnittlig z-skåre på 0,717. Riktig svar har også svært høy ”point-biserial” på hele 0,73.

Kodene for **Ikke riktig svar** og **Ikke svart** har alle sterk negativ gjennomsnittlig z-skåre og negativ ”point-biserial”. Under kode 70 var det tre elever som hadde ett av tallene feil på den tabellen som skulle fylles ut. Én av disse elevene svarte riktig på alle de andre spørsmålene i denne oppgaven.

Krysstabellen mellom spørsmål 21 og alle de andre spørsmålene i oppgaven viser at ingen under kode 71, 72 og 73 svarte riktig på noen av de andre spørsmålene i oppgaven. De besvarelsene som ble kodet 71 hadde misforstått hvordan tallfølgen fremkommer og ga tallfølgen 1, 1, 2, 4, 8, 16, 32, 64. De oppfattet spørsmålet dit hen at tallet 2 skulle opphøyes i eksponenter fra 2 til 6.

Under kode 72 var det 9 elever som oppga tallene 1, 1, 2, 3, 4, 5, 6, 7. Dette var svake elever, som den gjennomsnittlige z-skåren også viser. Antagelig ønsket de å oppgi et svar selv om de ikke oppfattet tallfølgen.

De tre besvarelsene under kode 73 oppga tallfølgen 11 22 33 44. Disse elevene svarte naturlig nok heller ikke riktig på de andre spørsmålene i denne oppgaven.

Under kodene for **Ikke riktig svar**, var det kode 79 som fikk den største tilslutningen. Dette var besvarelser det var vanskelig å finne et mønster i. Jeg antar at de fleste tippet en tallfølge, slik elevene under kode 72 og sannsynlig de under kode 73 nok også gjorde. Ser man på den gjennomsnittlige z-skåren, var dette i hovedsak svake elever, selv om to av disse fikk en liten positiv z-skåre. Det var også to elever som ikke svarte på spørsmålet, som også hadde positiv z-skåre.

Vi ser at 33 elever ikke besvarte spørsmålet. Ingen av disse besvarte de to påfølgende spørsmål heller. På det siste spørsmålet i oppgaven, spørsmål 24, var det tre av disse som oppga svar, hvorav to svarte riktig.

---

**Spørsmål 22:** FIBONACCITALLENE

Vi ønsker å fortsette med et nytt kvadrat H.

Hvor lange vil sidene i kvadrat H bli?

………..

---

## Koder til spørsmål 22

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 34 | 68 | 47,6 | 0,780 | 0,75 |
| | **Ikke riktig svar** | | | | |
| 70 | Følgefeil | 1 | 0,7 | -0,429 | -0,04 |
| 71 | 128 | 4 | 2,8 | -0,488 | -0,08 |
| 79 | Andre svar | 25 | 16,8 | -0,662 | -0,31 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 45 | 31,5 | -0,758 | -0,52 |

Intensjonen med spørsmålet var å registrere hvilke elever som oppfattet mønsteret for utviklingen av kvadrater med sider som har lengder fra Fibonaccitallfølgen. En krysstabell mellom spørsmål 21 og 22 viser at av de 72 elevene som svarte riktig på forrige spørsmål, var det 68 som svarte riktig her. Med frafallet av disse fire økte den gjennomsnittlige

z-skåren for kode 10 til 0,78. De fire hadde en gjennomsnittlig z-skåre på –0,341.

Tabellen viser at de andre svaralternativene både har negativ gjennomsnittlig z-skåre og ”point-biserial”. De 45 elevene som ikke svarte på spørsmålet har også den laveste gjennomsnittlige z-skåren, –0,758. Under denne koden er det tre elever som har z-skåre over null. Tilsvarende var det også for seks elever under **Ikke riktig svar**. Dette viser at det er flere elever med resultat over middels som ikke fikk tak i mønstret for utvikling av tallfølgen.

Jeg valgte å opprette kode 70 for den ene eleven som faktisk oppfattet mønstret, men kom ut med feil svar. Eleven summerte feil på spørsmål 21. Denne feilen ble videreført til dette spørsmålet. Det er et viktig prinsipp for de oppgaver som utvikles for PISA at de spørsmål som presenteres, ikke bør være avhengig av at eleven mestret tidligere spørsmål. Oppgavens hensikt var å introdusere en særegen tallfølge, og ved en slik oppgave var det vanskelig å forhindre denne svakheten. Det kan derved diskuteres hvorvidt oppgaven i tilstrekkelig grad tilfredsstilte kravene i en større internasjonal test som PISA-undersøkelsen.

De fire besvarelsene som ble kodet 71 oppga svaret 128. De fortsatte med det de oppfattet som en tallfølge utviklet ved potensen $2^N$ der $N$ nå var 7. En krysstabell mellom dette spørsmålet og de senere spørsmålene i oppgaven, viser at ingen av disse elevene oppfattet utviklingen av Fibonaccitallene.

---

**Spørsmål 23:** FIBONACCITALLENE

Dersom vi fortsetter å tegne slike kvadrater, vil vi etter hvert få et kvadrat med sider lik 89. Neste kvadrat vil da få sider lik 144.

Hvor lange sider får det neste kvadratet etter dette igjen?

………

Forklar hvordan du tenkte:

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

---

## Koder til spørsmål 23

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 233 med forklaring | 67 | 46,9 | 0,777 | 0,73 |
| 11 | 233 uten forklaring | 2 | 1,4 | -0,074 | -0,01 |
| | **Ikke riktig svar** | | | | |
| 70 | Feil regneprosedyre | 5 | 3,5 | -0,572 | -0,11 |
| 79 | Andre svar | 7 | 4,9 | -0,649 | -0,14 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 62 | 43,4 | -0,717 | -0,63 |

I utgangspunktet var det tenkt brukt to koder for riktig svar, kode 10 for de som hadde riktig svar uten forklaring, og kode 20 for de som også hadde med forklaring. Det ville føre til at det ble gitt to poeng på et spørsmål som i hovedsak ikke var mer arbeidskrevende enn å sette seg inn i en instruks som lett kunne hentes fra innledningen. Jeg valgte derfor i stedet å bruke kode 10 for riktig svar med forklaring og kode 11 for riktig svar uten forklaring. Elevenes forklaring på hvordan de tenkte, varierte sterkt fra de enkleste der de summerte tallene 89 og 144, til omfattende utledninger. Jeg valgte å kode alle disse under kode 10. Dersom oppgaven skulle ha andre funksjoner som for eksempel en diagnostisk test, ville det være av interesse å benytte flere koder for riktig svar og forklaring for nærmere å se hvordan den enkelte elev kom fram til riktig svar. Jeg vil også tillegge at kravet til forklaring til dette spørsmålet ikke burde vært med her, tatt i betraktning den opprinnelige intensjonen med denne testen og de få besvarelsene under kode 11. Krav om forklaring ville være mer formålstjenlig for andre spørsmål i denne testen for på den måten å finne ut hvilke metoder de har benyttet for å svare på spørsmålene.

Vanskegraden på spørsmålet holdt seg på nivå med de to første spørsmålene, selv om det var færre elever som besvarte spørsmålet. Det lå mer arbeid i besvarelsen på dette spørsmålet enn det foregående. På tross av lavere antall besvarelser, økte antall elever som svarte riktig med én (for de som ble kodet 10 og 11). Tidspress under testen kan være årsaken til at flere kandidater som svarte riktig på spørsmålet, ikke rettet opp gal besvarelse på det foregående. Det var fem elever som svarte riktig på dette spørsmålet, men ikke svarte riktig på foregående.

Besvarelsene under kode 11 fikk en liten negativ gjennomsnittlig z-skåre på

-0,074. Det var her to besvarelser, hvorav den ene besvarelsen trakk sterkt nedover. Krysstabell mellom dette spørsmålet og de andre i oppgaven viser at denne eleven ikke svarte riktig på noen av de andre spørsmålene. Det kan stilles spørsmål ved om eleven her fikk noe ”hjelp”. Kode 10 og kode 11 under ett gir litt laver gjennomsnittlig z-skåre på 0,752, mens ”point-biserial” ble omtrent den samme koeffisienten.

Under kode 70 var det fem besvarelser som alle oppga 199. Den forklaringen som ble gitt til svaret, var at de først fant differensen mellom 144 og 89 som de senere summerte med 144 slik at svaret ble 199. Det var kandidater fra flere av skolene som ga dette svaret. Tabellen viser at dette er meget svake elever med en gjennomsnittlig z-skåre og "point-biserial" på henholdsvis –0,571 og –0,11. Ingen av disse elevene svarte riktig på noen av de andre spørsmålene i oppgaven.

For de 7 besvarelsene som ble kodet 79, fant jeg ingen mønster bortsett fra at to elever skriver at de tippet et tall. Ikke uventet hadde de som ble kodet under 99 den laveste gjennomsnittlige z-skåren selv om seks av disse hadde positiv z-skåre.

---

**Spørsmål 24:** FIBONACCITALLENE

Sidelengdene for disse kvadratene utgjør den rekken av tall som vi altså kaller for Fibonaccitallene.

Vi kan sette opp deler av denne rekka slik:

1 , 1 , 2 , 3 , 5, … , … , … , … , … , 89 , 144 , … , … , … ,

Sett inn tallene som mangler ovenfor.

---

# Koder til spørsmål 24

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 8,13,21,34,55,…,377, 610 | 69 | 48,3 | 0,704 | 0,68 |
| | **Ikke riktig svar** | | | | |
| 70 | Følgefeil | 5 | 3,5 | 0,115 | 0,02 |
| 79 | Andre svar | 15 | 10,5 | -0,745 | -0,26 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 54 | 37,8 | -0,705 | -0,55 |

Elevene skulle ved dette spørsmålet sette inn de tall som manglet i tallfølgen. For de elevene som hadde oppdaget mønstret, var dette en oppsummering av de tre foregående spørsmål. For de andre, ville dette være en tankenøtt hvis de skulle ha noe tid.

P-verdien på dette spørsmålet ble i likhet med foregående på 0,48. Vi ser også av tabellen at de besvarelsene som er kodet 10 har noe lavere gjennomsnittlig z-skåre og "point-biserial" enn på de andre spørsmålene i oppgaven *Fibonaccitallene*.

En krysstabell mellom spørsmålene 23 og 24 viser årsaken.

**Krysstabell mellom spørsmål 23 og spørsmål 24**

Count

| | | Spørsmål 24 | | | | |
|---|---|---|---|---|---|---|
| | | 10 | 70 | 79 | 99 | Total |
| Spørsmål 23 | 10 | 64 | 3 | | | 67 |
| | 11 | 1 | | 1 | | 2 |
| | 70 | | | 3 | 2 | 5 |
| | 79 | 1 | | 5 | 1 | 7 |
| | 99 | 3 | 2 | 6 | 51 | 62 |
| Total | | 69 | 5 | 15 | 54 | 143 |

Tre elever som svarte riktig på spørsmål 23, ble på dette spørsmål kodet 70. Disse hadde en gjennomsnittlig z-skåre på 0,635. Samtidig var det også tre elever med en gjennomsnittlig z-skåre på –0,705 som ikke besvarte spørsmål 23, men som svarte riktig på spørsmål 24. Det er forunderlig hvordan de her fant tallfølgen når det også viser seg at to av disse elevene ikke besvarte de andre spørsmålene i oppgaven.

De fem elevbesvarelsene som ble kodet 70, ga et svar med små og for noen ubetydelige regnefeil på tallfølgen. Feilene gikk på unøyaktig summering av tall. Her var det også to kandidater med gjennomsnittlig z-skåre på –0,666, som ikke besvarte forrige spørsmål. Den gjennomsnittlige z-skåren for de andre tre besvarelsene ble på hele 0,635, ikke så mye lavere enn for besvarelsene under kode 10.

To av de fem besvarelsene som ble kodet 70 på spørsmål 24, ble kodet under **Ikke svart** på spørsmål 23. På spørsmål 24 presenterte de en tallfølge som ble avsluttet med regnefeil. De andre tre svarte også riktig på spørsmål 23, men endte også med regnefeil. To av disse svarte riktig på de tre andre spørsmålene i oppgaven og oppnådde dessuten høy z-skåre på testen.

Krysstabellen viser en sterk sammenheng mellom elevbesvarelsene på spørsmål 23 og 24. En Kji-kvadrattest basert på de som svarte riktig og de som ikke svarte riktig henholdsvis på det ene eller begge spørsmål, ga en verdi på hele 45,42. En tilsvarende test mellom de andre spørsmålene viser også entydig en klar sammenheng.

### *7.5.1 Oppsummering rundt oppgaven Fibonaccitallene*

Alle spørsmål i oppgaven oppnådde en p-verdi på omkring 0,5. En pilottest gjennomført i Australia (ACER) sommeren 2001 ga omtrent samme resultat. Med den ideelle p-verdien (og ikke minst den høye ”point-biserial”) skulle dette tyde på at spørsmålene isolert ville fungere bra i en testsammenheng på linje med det PISA ønsker av oppgaver.

Blant de 75 elevene som besvarte alle spørsmålene i oppgaven, var det så mange som 61 elever som svarte riktig på alle. Dette kan igjen tyde på at de fire spørsmålene i oppgaven var forholdsvis like og derfor målte mye av det samme. Antagelsen styrkes ved at alle spørsmålene i denne oppgaven korrelerte svært høyt med hverandre. Dessuten ga en ”point-biserial” korrelasjon mellom riktig besvarelse på de respektive spørsmålene og summen for alle fire spørsmål, en koeffisient fra 0,94 til 0,97. De meget høye verdiene på Kji-kvadrattesten for elevbesvarelsene, understreker dette. Den gjennomsnittlige z-skåren for de som svarte riktig på alle spørsmålene ble på 0,827 og "point-biserial" ble på hele 0,72, som viser klart at oppgaven *Fibonaccitallene* diskriminerte bra. Det var 30 elever som ikke svarte på noen av spørsmålene. Dette kan tyde på at flere elever styrer unna oppgaver med *Vitenskapelig* kontekst.

Denne oppgaven kom ikke med til Generalprøven. Dersom oppgaven skulle benyttes til PISA burde den gjennomgå en forandring ved at spørsmålene ble mer variert, eller eventuelt at noen av spørsmålene ble fjernet. Validiteten til et oppgavesett beror på hvilken funksjon oppgavene skal ha. Spørsmålene i oppgaven *Fibonaccitallene* slik den fremsto til pilottesten, har lav validitet i forhold til det som var ønsket til PISA-undersøkelsen. Alle fire spørsmålene måler mye av det samme. Det at hvert av spørsmålene korrelerer svært høyt med summen av riktig besvarelse for de fire spørsmålene, understreker dette.

Vi kan anta at fibonaccitallene er et nytt tema for elevene. Elevene skal ut fra en innledende tekst fulgt av noen spørsmål, selv komme fram til tallfølgen. Dersom testens hensikt er å måle elevenes dyktighet i å frembringe tallfølger, i dette tilfelle fibonaccitallene, anser jeg at oppgaven har høy validitet. De enkelte spørsmålene vil her fungere som en veiledning. Men det var ikke intensjonen da jeg utviklet oppgaven. Ved en annen bruk av oppgaven, kan det derved argumenteres for at den ville fungere utmerket.

## 7.6 Oppgaven Kast med to terninger

Oppgaven *Kast med to terninger* konsentrerer seg om sannsynlighetsregning, et forholdsvis nytt tema i norsk grunnskole. Innholdet i oppgaven vil det være naturlig å plassere i det rammeverket for PISA definerer som *Usikkerhet*, som er knyttet til statistikk og sannsynlighet.

Selv om mange elever viste stor grad av uklarhet rundt spørsmålene, har jeg på grunnlag av den matematikken de bruker vurdert spørsmål 25 til kompetanseklasse 1, og de tre siste til kompetanseklasse 2.

Situasjoner rundt spill med terningkast har de fleste elever stiftet bekjentskap med i oppveksten. Derfor vil problemstillingen rundt *Kast med to terninger* plasseres i distanse nærmest eleven, *Personlig liv*.

# KAST MED TO TERNINGER

Ved kast med to terninger, en rød og en blå, kan vi legge sammen antall prikker som vender opp på begge terningene.

Det er 5 forskjellige muligheter til å få summen 8 ved kast med begge terningene, for eksempel 3 på den røde terningen og 5 på den blå, eller 2 på den blå og 6 på den røde terningen.

I henhold til L97 bør elevene beherske følgende:

- *med utgangspunkt i praktiske erfaringer tilegne seg begreper om sannsynlighet (L97, s.166)*
- *arbeide med å utvikle mer presise begreper og uttrykksmåter for sannsynlighet og med å tallfeste sannsynligheter. (ibid, s.169)*
- *beregne sannsynligheter ut fra situasjoner hvor alle enkeltutfall har like stor sjanse. (ibid)*
- *undersøke situasjoner der det må regnes med usikkerhet, risiko og sjanse, f eks spill, forsikring, etterforskning og medisin. (ibid)*

**Spørsmål 25:** KAST MED TO TERNINGER

Hvor mange muligheter er det for å få summen 12?

………muligheter

## Koder til spørsmål 25

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 1 mulighet | 109 | 76,2 | 0,115 | 0,21 |
| | **Ikke riktig svar** | | | | |
| 70 | 2 muligheter | 14 | 9,8 | 0,323 | 0,11 |
| 71 | Ingen mulighet | 1 | 0,7 | 0,517 | 0,04 |
| 79 | Andre svar | 7 | 4,9 | -0,750 | -0,17 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 12 | 8,4 | -1,031 | -0,31 |

Det er oppmuntrende å se at så mange som 131 elever besvarte spørsmålet. Av disse var det 83,2 % som oppga riktig antall muligheter for å få summen 12 ved kast med to terninger. En grunn til at den gjennomsnittlige z-skåren for kode 10 bare ble på 0,115, var at mange elever som fikk til dels sterk negativ z-skåre på hele testen, her oppga riktig svar. Så mange som 58 (53,2 %) av de elevene som oppga riktig svar, fikk samlet negativ skåre for hele testen. Spørsmålet diskriminerte derfor meget svakt med ”point-biserial” på bare 0,21.

Av de elevbesvarelsene som ble kodet under **Ikke riktig svar** var det spesielt de 14 under kode 70 som er interessante å se nærmere på. De 14 besvarelsene oppga to muligheter for å få summen 12. Årsaken til svaret de oppga, kan være at den innledende teksten med presentasjon av to terninger, en rød og en blå, forvirret noen av elevene (mer om det under kommentarene til spørsmål 26).

Den gjennomsnittlige z-skåren for besvarelsene under kode 70 ble 0,323, noe høyere enn for de som ble kodet under kode 10. En grunn til det er at 8 av de 14 hadde en relativt høy z-skåre. Her igjen ser vi hvordan ”point-biserial”, som både tar hensyn til antallet og spredningen i utvalget, kommer til nytte. En krysstabell mellom spørsmål 25 og 26 (se under spørsmål 26) viser også at 10 av disse elevene svarte riktig på spørsmål 26 med en gjennomsnittlig z-skåre på hele 0,600. Dette tyder på en usikkerhet enkelte elever opplever når de skal regne ut kombinasjoner og sannsynlighet, spesielt med slik tekst oppgaven baserer seg på. Den ene besvarelsen under kode 71 som oppga null muligheter, hadde også en høy positiv z-skåre. Denne eleven svarte feil på alle spørsmålene i oppgaven *Kast med to terninger*.

Under koden 79 for andre ikke riktige svar var det én elev som svarte riktig både på spørsmål 26 og 28. Dette var den eneste eleven under denne koden, som hadde positiv z-skåre.

**Spørsmål 26:** KAST MED TO TERNINGER

Hvor mange muligheter er det for å få summen 7 ved kast med begge terningene?

……….muligheter

# Koder til spørsmål 26

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 6 muligheter | 51 | 35,7 | 0,556 | 0,42 |
| | **Ikke riktig svar** | | | | |
| 70 | 12 muligheter | 5 | 3,5 | 0,801 | 0,11 |
| 71 | 3 muligheter | 53 | 37,1 | -0,184 | -0,14 |
| 72 | 5 muligheter | 4 | 2,8 | -0,104 | -0,02 |
| 79 | Andre svar | 15 | 10,5 | -0,374 | -0,139 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 15 | 10,5 | -1,108 | -0,38 |

På spørsmålet skulle elevene bare oppgi antall muligheter det er for å få summen 7 ved kast med begge terningene. Da er det naturlig å ha bare én kode for riktig svar siden det ikke var krav om forklaring. I ettertid ser jeg at spørsmålet også burde bedt om forklaring slik at elevene fikk anledning til å begrunne sine svar. Forklaringen ville eventuelt avdekke om den innledende teksten med forskjellige farger på terningene, forvirret noen elever.

Spørsmålet fikk mye lavere p-verdi enn det foregående spørsmålet selv om svarprosenten var ubetydelig lavere. Til gjengjeld viser gjennomsnittlig z-skåre for kode 10 at det, i motsetning til spørsmål 25, i hovedsak var de flinkere elevene som svarte riktig, selv om 15 elever hadde negativ z-skåre på testen. Jeg anser at verdien for ”point-biserial” viser at spørsmålet diskriminerte betraktelig bedre enn det foregående spørsmål, selv om den høye gjennomsnittlige z-skåren for kode 70 som vi også så på det foregående spørsmål, skjemmer noe.

Det er fire koder for **Ikke riktig svar** der kode 70 skiller seg ut med høy gjennomsnittlig z-skåre på hele 0,801, vesentlig høyere enn gjennomsnittet for kode 10. Her får vi igjen nytte av ”point-biserial” som er på 0,15. De fem besvarelsene under denne koden hadde z-skåre fra

–1,257 til 2,173. Svaret de oppga var 12 muligheter for å få summen 7 ved kast av begge terningene. Her igjen er det et eksempel på tolkningsproblemer ved den innledende teksten som oppga forskjellige farger på terningene. Eksemplet i innledningen var gitt for å forklare at det var fem forskjellige muligheter for å få summen 8. Hvis elevene hadde forholdt seg til eksemplet ved selv å kontrollregne, ville trolig noen av dem oppgitt riktig svar på dette spørsmålet. Vi ser også av krysstabellen at fire av de fem under denne koden, svarte riktig på foregående spørsmål.

Det var hele 53 besvarelser under kode 71 som oppga tre muligheter. De halverte antall utfall ved ikke å ta hensyn til hvilken terning som viste antall prikker. Min intensjon med å gi terningene farge var nettopp for å forhindre dette. Ser man på z-skåren disse oppnådde på hele testen, er dette en meget sammensatt gruppe. 16 hadde en skåre over null, og av disse var det 6 med z-skåre på over 0,850. Dette viser klart den store usikkerhet også mange flinke elever opplever ved sider ved sannsynlighetsregning. Dette gjelder ikke minst ved tolkning av en oppgavetekst.

De fire elevene under kode 72 oppga 5 muligheter. De svarte alle riktig på spørsmål 25 (se krysstabellen under). Siden spørsmålet ikke krevde forklaring, er det umulig å fastslå hvilket utfall disse elevene ikke tok med.

Under kode 79 var det fem elever med positiv z-skåre, mens alle de som ikke besvarte spørsmålet hadde negativ skåre.

En krysstabell for spørsmål 25 og spørsmål 26 viser følgende:

**Krysstabell for spørsmål 25 og spørsmål 26**

Count

| | | Spørsmål 26 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 10 | 70 | 71 | 72 | 79 | 99 | Total |
| Spørsmål 25 | 10 | 40 | 4 | 49 | 4 | 9 | 3 | 109 |
| | 70 | 10 | 1 | 3 | | | | 14 |
| | 71 | | | 1 | | | | 1 |
| | 79 | 1 | | | | 5 | 1 | 7 |
| | 99 | | | | | 1 | 11 | 12 |
| Total | | 51 | 5 | 53 | 4 | 15 | 15 | 143 |

Det var 40 elever som svarte riktig på begge spørsmålene. Disse hadde en gjennomsnittlig z-skåre på 0,555, omtrent den samme som de som svarte riktig på spørsmål 26. "Point-biserial" for disse 40 elevene ble på 0,35, noe lavere, som skyldes blant annet at antallet var lavere. Andelen elever med negativ z-skåre ble omtrent den samme som på spørsmålet 26 (30 %). Vi ser også at så mange som 66 elever som svarte riktig på spørsmål 25, ikke oppga riktig antall muligheter for å få summen 7 ved kast med begge terningene. Her er det ikke bare tolkningsproblemer med den innledende teksten som er årsaken. Den største gruppen, de 49 elevene som er kodet 71 på spørsmål 26 og kodet 10 på spørsmål 25, tok ikke hensyn til hvilken terning som viste hvilket antall prikker, som innledningen så klart ga et eksempel på. Over 30 % av disse fikk positiv z-skåre på hele testen.

**Spørsmål 27:** KAST MED TO TERNINGER

Sannsynligheten for å få en sekser ved kast av en terning er 1/6.

Hvor stor er da sannsynligheten for å få summen 12 ved kast med begge terningene?

………..

## Koder til spørsmål 27

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 1/36 | 17 | 11,9 | 1,171 | 0,43 |
| | **Ikke riktig svar** | | | | |
| 70 | 1/12 | 48 | 33,6 | 0,073 | 0,05 |
| 71 | 1/6 | 34 | 23,8 | -0,033 | -0,02 |
| 79 | Andre svar | 22 | 15,4 | -0,080 | -0,03 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 22 | 15,4 | -0,935 | -0,40 |

Det er flere mulige måter å komme frem til riktig svar. For de elever som er kjent med at et kast med to terninger gir 36 muligheter og bare en av disse gir to seksere, vil det være enkelt å anta at sannsynligheten blir 1/36. Det kan også tenkes at noen benyttet produktsetningen for uavhengige hendelser. Sannsynligheten for å få en sekser på en terning er 1/6. To seksere på to terninger gir i følge produktsetningen 1/36. Det var nedslående å se at bare 12 % av elevene svarte riktig.

Den svært høye gjennomsnittlige z-skåren viser at elevene som oppga riktig svar, var blant de flinkeste. En frekvenstabell på z-skårene viser også at 15 av de 17 hadde positiv z-skåre på hele testen. På spørsmål 25 ble det spurt etter antall muligheter for å få summen 12. Det var 16 av de 17 elevene som svarte riktig på spørsmål 27 som også svarte riktig på spørsmål 25. Den siste eleven oppga to muligheter der. Siden dette var en elev med z-skåre på 1,937 er det rart at eleven ikke gikk tilbake og rettet opp svaret på spørsmål 25, dersom ikke eleven på dette spørsmålet benyttet produktsetningen.

Tatt i betraktning det lave antallet som svarte riktig, ble ”point-biserial” for kode 10 svært høy. Dette viser at spørsmålet diskriminerte godt mellom de aller beste og de øvrige, selv om det var en stor gruppe kodet 70 under **Ikke riktig svar** som hadde en positiv gjennomsnittlig z-skåre, om ikke like stor.

De 48 besvarelsene som ble kodet 70 oppga alle sannsynligheten 1/12. 50 % av disse hadde positiv z-skåre på hele testen. Det er vanskelig å fastslå hvordan de endte opp med dette svaret. En krysstabell mellom spørsmål 27 og spørsmål 25 viser at 40 av disse svarte riktig på spørsmål 25. De var da kjent med at det var én mulighet for å få summen 12 ved kast med begge terningene. Problemet for disse elevene var antagelig utfallsrommet. En terning gir som kjent 6 mulige utfall. De har trolig oppfattet at to terninger bør gi 12 mulige utfall, og ikke 36.

Det blir tilsvarende spekulasjoner omkring hvordan elevene kom fram til tallet 1/6 for de som ble kodet 71. 24 av de 34 som kom under denne koden hadde svart riktig på spørsmål 25. Det er å anta at flere av disse hadde en oppfatning om at sannsynligheten blir den samme ved kast med to terninger som med en. To av elevene under kode 79 oppga sannsynligheten 2/6. Antagelig summerte de isteden for å multiplisere. De hadde begge svært høy z-skåre på testen.

En krysstabell mellom kjønn og spørsmålet viser en interessant situasjon.

**Krysstabell mellom kjønn og spørsmål 27**

Count

| | | Spørsmål 27 | | | | | |
|---|---|---|---|---|---|---|---|
| | | 10 | 70 | 71 | 79 | 99 | Total |
| kjønn | Jenter | 5 | 28 | 22 | 11 | 13 | 79 |
| | Gutter | 12 | 20 | 12 | 11 | 9 | 64 |
| Total | | 17 | 48 | 34 | 22 | 22 | 143 |

En Kji-kvadrattest med verdien 5,21 viser en klar sammenheng mellom kjønn og riktig svar på spørsmålet. Dette var det eneste spørsmålet i pilottesten der guttene gjorde det signifikant bedre enn jentene. Denne forskjellen kan ikke være forårsaket av at den innledende teksten til oppgaven forvirret noen elever. På de andre spørsmålene i oppgaven ble det nemlig ikke registrert slike forskjeller.

---

**Spørsmål 28:** KAST MED TO TERNINGER

Hvor stor er sannsynligheten for å få summen 7 ved kast med begge terningene?

Sett en ring rundt det svaret du mener er riktig.

A 7/36

B 1/12

C 1/6

D 1/36

---

## Koder til spørsmål 28

Riktig svar er her merket med*.

| **Koder** | **Frekvens** | **Prosent** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| A | 33 | 23,1 | -0,204 | -0,11 |
| B | 28 | 19,6 | 0,044 | 0,02 |
| C* | 32 | 22,4 | 0,499 | 0,27 |
| D | 23 | 16,1 | 0,106 | 0,05 |
| **Ikke svart** | | | | |
| 99 | 27 | 18,9 | -0,478 | -0,23 |

Denne flervalgsoppgaven dokumenterte den usikkerhet mange elever opplever med sannsynlighetsregning. Antall elever som ikke besvarte spørsmålet har gjennom oppgaven Kast med to terninger sakte økt fra 8,4 % til 18,9 %. Sett på en annen måte er jeg rimelig fornøyd med at så mange som 81 % besvarte dette spørsmålet, selv om antagelig mange av disse følte en usikkerhet i forhold til det alternativet de valgte. Den store tilslutning de forskjellige distraktorene fikk, viser også at mange var usikre på alternativene. Spørsmålet ville tjent på at elevene måtte begrunne valg av alternativ. På den måten ville de forhåpentlig i større grad bli tvunget til å overveie alternativene. Dessuten ville det gitt verdifull informasjon om hvordan de tenkte, men samtidig ville det også være en fare for at færre ville besvare spørsmålet. På dette spørsmålet har mange opplagt tippet på alternativene.

Selv om p-verdien ble betraktelig høyere enn på foregående spørsmål, viser både den gjennomsnittlige z-skåren og "point-biserial" at det også var flere svake elever som her hadde riktig svar. En frekvenstabell på z-skåren for riktig svaralternativ viser at 34,4 % hadde negativ skåre, mot bare 11,8 % på foregående spørsmål.

Tabellen viser også at distraktoren A fikk større tilslutning enn det riktige alternativet. Den hadde også den laveste gjennomsnittlige z-skåren av alle alternativene. Av de som er kodet under denne distraktoren, var det 39,4 % som hadde positiv z-skåre på hele testen, for noen av dem var den meget høy. Tilsvarende var det hele 52,2 % som hadde positiv z-skåre på distraktor D. Det var ingen elever som valgte denne distraktoren, som svarte riktig på både spørsmål 26 og spørsmål 27.

Det vil være urimelig å anta at den store oppslutningen til distraktorene skyldtes at de var spesielt gode, men heller at mange tok en sjanse og valgte den de syntes passet best uten selv å regne etter. Ved ren tipping ville det være 25 % sjanse å velge riktig, mens resultatet her var at 27,6 % av de som svarte, krysset av for riktig alternativ. Ikke langt fra ren tipping. Forskjellen er for liten til å påstå at det er signifikant forskjell på resultatet og hva som ville fremkomme ved bare tipping. En test vil gi en signifikantsanssynlighet på hele 0,44 (se punkt 4.1.2 og 4.1.3). Dersom vi setter signifikansnivået på 0,05, burde det være minst 43 elever som krysset av riktig blant de som besvarte oppgaven.

Det er naturlig å håpe at de som svarte riktig på spørsmålet, også mestret foregående spørsmål, siden de på begge baserte seg på et utfallsrom på 36 som vi har ved kast av to terninger. En krysstabell for spørsmålene viser følgende:

**Krysstabell mellom spørsmål 27 og spørsmål 28**

Count

| | | Spørsmål 28 | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3* | 4 | 99 | |
| Spørsmål 27 | 10 | 4 | 4 | 7 | 1 | 1 | 17 |
| | 70 | 13 | 15 | 9 | 5 | 6 | 48 |
| | 71 | 12 | 3 | 7 | 7 | 5 | 34 |
| | 79 | 3 | 5 | 7 | 7 | | 22 |
| | 99 | 1 | 1 | 2 | 3 | 15 | 22 |
| Total | | 33 | 28 | 32 | 23 | 27 | 143 |

Her ser vi at av de 17 som svarte riktig på spørsmål 27, var det 7 elever som svarte riktig på begge spørsmålene. De var svært flinke med en gjennomsnittlig z-skåre på hele 1,794, og ingen hadde lavere skåre enn 1,109. De andre kombinasjonene mellom riktig svar på spørsmål 28 og **Ikke riktig svar** på foregående spørsmål, gir en frekvenstabell med stor spredning av z-skårene, og med et gjennomsnitt noe over null.

På spørsmål 26 ble det spurt om antall muligheter det er for å få summen 7 og på spørsmål 28 sannsynligheten for å få summen 7. Det er naturlig å se på svarmønstret for begge spørsmålene. Krysstabellen viser følgende:

**Krysstabell mellom spørsmål 26 og spørsmål 28**

Count

| | | Spørsmål 28 | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3* | 4 | 99 | |
| Spørsmål 26 | 10 | 8 | 7 | 17 | 12 | 7 | 51 |
| | 70 | 1 | 4 | | | | 5 |
| | 71 | 18 | 12 | 9 | 7 | 7 | 53 |
| | 72 | | 2 | 1 | 1 | | 4 |
| | 79 | 4 | 3 | 5 | 2 | 1 | 15 |
| | 99 | 2 | | | 1 | 12 | 15 |
| Total | | 33 | 28 | 32 | 23 | 27 | 143 |

Her ser vi at 17 elever svarer riktig på begge. De har en gjennomsnittlig z-skåre på 0,962, der 14 skårer positivt. Samtidig ser vi at det var 54 elever som svarte feil på begge. En Kji-kvadrattest for alle som besvarte begge spørsmålene viste imidlertid at det var liten signifikant sammenheng mellom besvarelsene.

Det var 8 elever som valgte distraktor A og svarte riktig på spørsmål 26. Ingen av disse svarte riktig på spørsmål 27. Dette var i hovedsak svake elever med gjennomsnittlig z-skåre på –0,341.

De 28 besvarelsene som valgte distraktor B, hadde en liten positiv gjennomsnittlig z-skåre. 15 av disse hadde også svart 1/12 på foregående spørsmål. Det er naturlig å anta at flere la til grunn samme resonnement som de gjorde på spørsmål 27 (se kommentarer for kode 70). Av de som valgte distraktor B, var det én elev som svarte riktig på både spørsmål 26 og spørsmål 27. Dette var en flink elev med z-skåre på 0,517 på hele testen.

Elevene som valgte distraktor D, hadde også en liten positiv gjennomsnittlig z-skåre. Her var det ingen elever som samtidig svarte riktig på begge de to foregående spørsmål. Det var 12 elever som svarte riktig på spørsmål 26, og det var bare én elev som svarte riktig på spørsmål 27. Denne eleven viste der at to terninger gir 36 mulige enkeltutfall, men tok ikke med antall muligheter for å få summen 7. De 12 elevene som hadde oppgitt en sannsynlighet på enten 1/12 eller 1/6 for å få summen 12 ved kast av to terninger på foregående oppgave, valgte her 1/36 for å få summen 7. Det kan tyde på at de tippet alternativet på dette spørsmålet.

### 7.6.1 Oppsummering rundt oppgaven Kast med to terninger

Det var bare 6 elever som svarte riktig på alle fire spørsmål i denne oppgaven. De hadde en meget høy gjennomsnittlig z-skåre på hele 1,759. Disse besvarelsene korrelerte bra med testen som helhet tatt i betraktning de få elevene som svarte riktig, selv om koeffisientverdien for "point-biserial" ble på 0,37. Med 6 riktige besvarelser gir da oppgaven en p-verdi på 0,04. Denne lave p-verdien vil selvfølgelig påvirke "point-biserial" korrelasjonen (se formelen under p.4.1.4). Hvis denne oppgavens hensikt var å registrere de få elevene som behersket sannsynlighetsregning i denne oppgaven bra, diskriminerte den tilfredsstillende.

Selv om det var få elever som besvarte alle fire spørsmålene riktig, er det oppmuntrende å se at så mange som 107 elever (74,8 %) besvarte alle spørsmålene, og at det bare var ni elever (6,3 %) som ikke besvarte noen av spørsmålene. Disse ni var blant de svakeste elevene, noe deres gjennomsnittlige z-skåren på –1,205 også viser. Den høye svarprosenten kan tyde på at flere elever "tenner" mer på oppgaver med problemstillinger hentet i distanse nær elevene, *Personlig liv*.

Med de resultater pilottesten ga, har jeg forståelse for at denne oppgaven ikke kom med til Generalprøven i PISA. Ved siden av høy p-verdi på spørsmål 25, diskriminerte den meget svakt mellom besvarelsene under **Riktig svar** og de under **Ikke riktig svar**. Selv om spørsmål 26 hadde en akseptabel p-verdi, viste krysstabellen at bare 36,7 % av de som svarte riktig på dette spørsmålet også svarte riktig på foregående. P-verdien på spørsmål 27 ble på bare 0,12, mot 0,76 på spørsmål 25. Det første spørsmålet spurte om mulighetene for å få summen 12, mens det andre spurte om sannsynligheten for å få summen 12. Dette viser klart den usikkerhet mange elever har rundt sannsynlighetsregning.

Det siste spørsmålet i oppgaven diskriminerte også meget svakt, med stor tilslutning til distraktorene også av de som hadde positiv z-skåre på hele testen.

Selv om oppgaven ville være lite velegnet til PISA-undersøkelsen eller en tradisjonell prøve, ser jeg andre verdier ved oppgaven. Den er meget egnet til å identifisere misoppfatninger som elevene har utviklet ved sannsynlighetsregning. Senere har jeg hatt stort utbytte av denne oppgaven i min undervisning. Etter en kort introduksjon av sannsynlighetsregning har jeg gitt elevene en omarbeidet utgave av oppgaven. De skulle i tillegg til å gi svar, også begrunne hvordan de kom fram til disse svarene. Dette har gitt meg viktig informasjon om de løsningsstrategier de har benyttet, og har gitt meg et grunnlag for å planlegge undervisningsopplegget i sannsynlighetsregning tilpasset den enkelte elev.

På samme måte som å følge svarmønstre ved hjelp av krysstabeller for derigjennom å få kunnskaper for eventuelt å korrigere spørsmål i en oppgave, vil altså den samme fremgangsmåten også gi viktig informasjon om den enkelte elevs løsningsstrategier. På den måten kan det observeres misoppfatninger elevene har utviklet. Derfor anser jeg at oppgaven vil fungere bedre som en diagnostisk oppgave enn til å kartlegge elevenes dyktighet i sannsynlighetsregning.

## 7.7 Oppgaven Terninger

I Fengselsundervisningen møter jeg ofte elever som ikke har funnet seg til rette i ordinær skole. De fleste har reserverte holdninger til matematikk, som de har opplevd både som vanskelig og lite engasjerende. Til disse elevene har jeg utviklet oppgaver som viser andre sider av matematikken enn tall og algebra, behandling av data og grafer og funksjoner. Ofte starter jeg undervisningen med figurer i planet og rommet. Oppgaven *Terninger* er en revidert utgave av slike oppgaver.

Konteksten til denne oppgaven er i likhet med oppgavene *Potete*r og *Kast med to terninger*, hentet fra distanse nærmest eleven, *Personlig liv*. Det var da naturlig å forvente en relativt høy svarprosent i likhet med for de to andre oppgavene. Det første spørsmålet i oppgaven er fra kompetanseklasse 1, spørsmålene 30, 31 og 33 er fra kompetanseklasse 2, og spørsmål 32 er fra kompetanseklasse 3.

Læreplanverket for den 10-årige grunnskolens småskoletrinn gir inntrykk av at målområdet Rom og form er en introduksjon til geometrien elevene møter senere (L97, s.156). I like høy grad vil dette målområdet også gi elevene utfordringer til å finne mønstre og oppdage strukturer i enkle og sammensatte figurer. Dette er et område som er kommet inn i den nye læreplanen for den videregående skolen som gjelder fra år 2000, ved *Eksperimentell geometri* i 1MY og *Geometri med polyedre* i 2MZ.  Det er naturlig å sette alle spørsmålene i oppgaven under det som PISA definerer som *Rom og form*.

I de tre første spørsmålene skal elevene ved hjelp av regelen for konstruksjon av terning, forholde seg til hvilke sider som står mot hverandre. I de to siste spørsmålene er utfordringen hvilke sider som forholder seg til hverandre når et pappstykke klippes til for så å brettes sammen til en terning.

# TERNINGER

Terningene som vi bruker i dag er en videreutvikling av den tidligere brukte *astragalus*, et lite benstykke fra foten til en sau eller geit. Den var avlang, og endestykket var avrundet slik at man bare kunne bruke fire av sidene. Disse sidene var ikke helt like, noe som førte til at det ikke var samme sjanse for hver av sidene.

Etter hvert ble eiendomsforhold og andre stridsspørsmål avgjort ved terningkast. Dette tvang fram behovet for en terning med samme sjanse for alle sidene, og de to endestykkene ble nå også tatt i bruk. Derved fikk vi dagens terning med seks sider.

En regel ved konstruksjon av dagens terning er at summen av antall prikker på to motstående sider alltid er syv.

---

Allerede tidlig i grunnskolen legges det vekt på at elevene skal

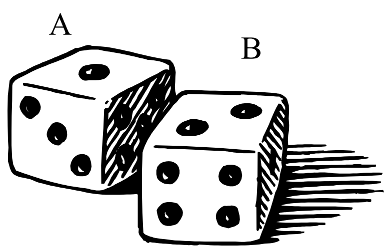
- *utvikle sine kunnskaper om rom og form og på den måten stimulere sin kreativitet og fantasi (L97, s.158).*
- *bli kjent med forskjellige figurer, former og mønstre og bruken av dem av dem i dagligliv og nærmiljø (ibid).*
- *lære å beskrive plassering i plan og rom (ibid, s.162).*
- *bli kjent med avbildninger, bruken av dem og sammenhengen med symmetri, mønster, modeller, kart og arbeidstegninger (ibid).*
- *lage figurer, former og mønstre, og arbeide med å finne ut av egenskaper ved dem (ibid, s.163).*

---

**Spørsmål 29:** TERNINGER

På figuren til høyre er det to terninger.

Hvor mange prikker vender ned på disse terningene?




Terning A:……..prikker.

Terning B:……..prikker.

## Koder til spørsmål 29

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 6 og 5 | 97 | 67,8 | 0,410 | 0,60 |
| | **Ikke riktig svar** | | | | |
| 70 | Riktig på den ene | 23 | 16,1 | -0,650 | -0,29 |
| 79 | Andre svar | 7 | 4,9 | -0,903 | -0,21 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 16 | 11,2 | -1,154 | -0,41 |

En av hensiktene med spørsmålet var å teste hvorvidt elevene hadde fått med seg regelen for konstruksjon av terninger slik den ble presentert i innledningen. Elevene fikk dermed en forberedelse på de to neste spørsmålene i oppgaven.

En svarprosent på 89 og p-verdi 0,68 ville normalt være tilfredsstillende. Spørsmålet diskriminerte tilfredsstillende med en "point-biserial" på 0,60 og kun kode 10 hadde positiv gjennomsnittlig z-skåre. Likevel er jeg noe forundret over at så mange som 30 elever oppga feil svar på et så enkelt spørsmål. Problemstillingen var å finne differensen mellom 7 og det antall prikker hver av terningene viste. Det er også forunderlig at så mange som 23 elever (besvarelsene under kode 70), hadde riktig på den ene terningen og ikke den andre. De fleste av disse måtte ha forstått regelen for konstruksjon av spillterninger, men bare to mestret spørsmål 30 og ingen spørsmål 31 (se krysstabell under spørsmål 30).

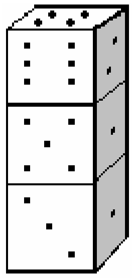
De 7 besvarelsene kodet 79 oppga feil antall prikker på begge terningene, men én av disse svarte riktig på spørsmål 30. Dette var en svak elev med en z-skåre på –0,666 for hele testen. De 16 elevbesvarelsene som ble kodet 99 hadde alle negativ z-skåre på testen. De fikk følge av ytterligere 30 på neste spørsmål. Ingen av de 16 svarte riktig på spørsmål 31, som også forholdt seg til regelen "*summen av antall prikker på to motstående sider er alltid syv*".

---

**Spørsmål 30:** TERNINGER

Vi stabler tre terninger over hverandre og ser at den øverste siden har fire prikker på den flaten som vender opp.

Hva er summen av prikkene på de fem skjulte vannrette sidene?



……….prikker

## Koder til spørsmål 30

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 17 | 60 | 42 | 0,698 | 0,60 |
| | **Ikke riktig svar** | | | | |
| 70 | Summeringsfeil | 1 | 0,7 | 0,990 | 0,08 |
| 71 | 21 (Inkl. topp) | 2 | 1,4 | 0,694 | 0,08 |
| 72 | 24 (skjulte loddr.) | 4 | 2,8 | 0,340 | 0,06 |
| 73 | 16, 18, 19 | 12 | 8,4 | -0,055 | -0,027 |
| 74 | 13, 14, 15 | 5 | 3,5 | -0,500 | -0,10 |
| 79 | Andre svar | 13 | 9,1 | -0,693 | -0,22 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 46 | 32,2 | -0,728 | -0,50 |

Både svarprosenten og p-verdien ble her lavere enn på foregående spørsmål. Den gjennomsnittlige z-skåren økte, noe som viser at andelen flinke elever her var høyere. "Point-biserial" for riktig svar ble praktisk talt den samme.

Det var mulig å benytte flere framgangsmåter for å komme frem til riktig svar. På enkelte av besvarelsene var det lett å se hvilken strategi som var benyttet. Noen elever benyttet en tungvint framgangsmåte ved å notere ned antall prikker på de skjulte loddrette sidene etter konstruksjonsregelen. På den måten fant de hvilke sider som var igjen på terningene, og dermed antall prikker på de skjulte vannrette sidene. Andre elever unngikk denne litt tungvinte framgangsmåten. De la i stedet til grunn at motstående sider på en terning, i dette tilfelle de vannrette sidene, alltid viser til sammen 7 prikker. Den tredje øverste terningen viser fire prikker på den øverste vannrette siden. Den skjulte vannrette siden har derfor tre prikker. Summen blir derved 17. En krysstabell mellom spørsmålene 29, 30 og 31 (se under spørsmål 31) viser at tre elever som mestret spørsmål 30, ikke mestret spørsmål 29 som var langt mindre utfordrende.

Det var flere av kodene under **Ikke riktig svar** som fikk positiv gjennomsnittlig z-skåre. Jeg valgte å opprette kode 70 for den eleven som oppga 9 prikker. I følge de notater som ble skrevet i besvarelsen, resonnerte eleven riktig i forhold til konstruksjonsregelen. Eleven noterte riktig antall prikker på de skjulte sidene på figuren, men oppga feil svar. Det må også legges til at eleven skåret bra på hele testen, svarte riktig både spørsmålene 29 og 31 og delvis riktig på spørsmål 32.

De to besvarelsene under kode 71 benyttet konstruksjonsregelen, men hadde inkludert den synlige vannrette siden og oppga derfor 21 prikker. De hadde en z-skåre nær opp til besvarelsene under kode 10.

Ved nærmere betraktning av besvarelsene under kode 72, kom det fram at de hadde forvekslet vannrette og loddrette sider. Derfor oppga disse elevene 24 prikkene på de loddrette sidene som ikke vises på tegningen. På et tidligere utkast til dette spørsmålet brukte jeg formuleringen de *fem skjulte horisontale sidene*. En test jeg gjennomførte på mine elever, viste at flere var usikre på begrepet *horisontale sider*. Jeg valgte derfor å benytte *vannrette sider*. Pilottesten viste at også ved bruk av dette uttrykket var det noen elever som fortsatt var usikre.

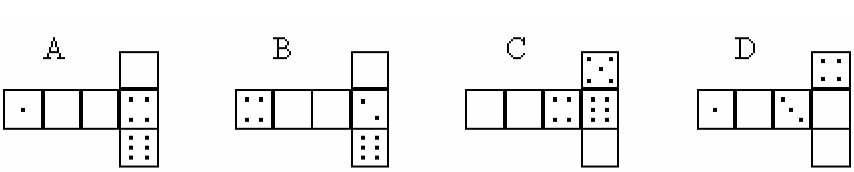
De 12 besvarelsene under kode 73 og de 5 besvarelsene under kode 74 hadde enten summert feil eller tippet feil antall prikker. Hvilket av alternativene som var grunnlaget for galt svar var vanskelig å fastslå siden de ikke hadde gjort notater på tegningen. Jeg valgte å opprette to forkjellige koder for disse besvarelsene for nærmere å se hvordan de svarte på de to andre spørsmålene i oppgaven som baserte seg på reglene for konstruksjon av terning til spill. Besvarelsene under kode 73 oppga enten 16 eller 18 prikker. De under kode 74 hadde et avvik på to prikker. Den gjennomsnittlige z-skåren for elevene under de respektive koder viser en klar forskjell. En frekvenstabell for z-skåren for de under kode 73 viser at noen flinke elever er kodet der. Krysstabell mellom spørsmålene 29, 30 og 31 viser at det var fire elever her som svarte riktig på spørsmålene 29 og 31 (se under spørsmål 31). Disse fire hadde en gjennomsnittlig z-skåre på hele 1,109. På de to andre spørsmålene viste de at de forsto regelen for konstruksjon av terninger. Derfor er det å anta at de på dette spørsmålet hadde summert feil.

---

**Spørsmål 31:** TERNINGER

Det er flere måter å klippe ut et sammenhengende pappstykke, brette og lime sammen for å lage en terning. Her er fire slike "utbrettede" terninger, og det er avsatt prikker på tre av sidene.

Sett en ring rundt den/de figurene du tror kan bli en terning som følger reglene om konstruksjon av terning til spill.




---

Ved dette spørsmålet får elevene igjen prøve seg på konstruksjonsreglen for terning til spill. I tillegg er nå terningene brettet ut slik at elevene også må kunne se hvilke sider som vil stå i forhold til hverandre når pappstykkene blir brettet og limt sammen.

## Koder til spørsmål 31

| Koder | Frekvens | Prosent | Prosent av de som svarte | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig** | | | | |
| 1 | 52 | 36,4 | 45,2 | 0,797 | 0,60 |
| | **Ikke riktig** | | | | |
| 0 | 63 | 44,1 | 54,8 | -0,245 | -0,21 |
| | **Ikke svart** | | | | |
| 99 | 28 | 19,6 | | -0,928 | -0,46 |

En stor mangel ved dette spørsmålet slik det her ble presentert, var at det ikke var med en tabell under der elevene skulle sette en ring rundt **Ja**/**Nei** for hvert utklipp. I min retting av pilottesten var det ikke mulig å registrere forskjellen på **Nei** (kan ikke bli en terning som følger regelen), og der eleven unnlot å svare.

Dette ble senere rettet opp til Generalprøven. Derved ble de figurene som ble satt en ring rundt registrert som enten Ja eller Nei. Med dette økte reliabiliteten betraktelig.

De 28 elevene kodet 99 som ikke svarte på spørsmålet, var de som ikke satte ring rundt noen av figurene. Tatt i betraktning av at oppgaven dukket opp sent i testen, er jeg rimelig fornøyd med at over 80 % oppga et svar. Av disse var det 44,1 % som valgte ut de riktige terningene B og C. Den akseptable p-verdien og "point-biserial" på 0,60 skulle tilsi at spørsmålet diskriminerte bra. En frekvenstabell for z-skårene for de som mestret spørsmålet, viser at mange svært flinke elever fikk den til. Samtidig falt den i smak også hos noen svake elever. Et eksempel er en elev som oppnådde bare 4 riktige svar totalt på hele testen, hvorav to under denne oppgaven, blant annet dette spørsmålet.

Av de 28 elevene som ikke besvarte spørsmålet, var det bare én med positiv z-skåre. Eleven svarte riktig på de to foregående spørsmål. Eleven kom antagelig i tidsnød. De siste spørsmålene i testen ble ikke besvart.

En tabell over besvarelsene til de enkelte figurene viser at riktig alternativ fikk desidert størst tilslutning. Her er ikke de 28 elevene som ikke besvarte spørsmålet tatt med. Dette var i hovedsak svake elever med en gjennomsnittlig z-skåre på - 0,928. Bare én av disse fikk positiv z-skåre på hele testen.

## Koder til de respektive figurer i spørsmål 31

I tabellen står kode **Ja** for *kan bli en terning*, dvs. at eleven har satt en ring rundt dette alternativet, og kode **Nei** for *kan ikke bli en terning*. Dvs. at eleven ikke har satt ring rundt dette alternativet. Derved er de som har unnlatt å ta stilling og derfor ikke satt ring, kodet som nei. Riktig svar er merket med *.

| **Koder til de respektive figurer** | **Frekvens** | **Prosent av de som svarte** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| **Figur A** | | | | |
| Ja | 14 | 12,2 | -0,150 | -0,05 |
| Nei* | 101 | 87,8 | 0,278 | 0,43 |
| **Figur B** | | | | |
| Ja* | 81 | 70,4 | 0,418 | 0,48 |
| Nei | 34 | 29,6 | -0,231 | -0,13 |
| **Figur C** | | | | |
| Ja* | 77 | 67 | 0,546 | 0,59 |
| Nei | 38 | 33 | -0,423 | -0,26 |
| **Figur D** | | | | |
| Ja | 14 | 12,2 | -0,277 | -0,09 |
| Nei* | 101 | 87,8 | 0,296 | 0,46 |

Tilsynelatende kommer det klart frem av tabellen at de figurene som kan foldes sammen til en terning i følge reglene, altså figurene 31B og 31C, fikk lavest prosentvis tilslutning for riktig svar med henholdsvis 70,4 % og 67 % riktig av de som besvarte spørsmålet. Årsaken kan ligge i mangelen på en etterfølgende tabell der elevene skulle angi sitt svar. De to andre figurene som ikke kunne foldes sammen til en terning i følge reglene, fikk begge 87,8 % riktig. Det var midlertidig 10 elever som krysset av feil på den ene og ikke den andre. Det var fire som krysset av feil på begge disse figurene.

Hvis spørsmålet ble presentert med en tabell der elevene skulle velge **Ja**/**Nei** for hvert av utklippene vil jeg anta at andelen som svarte feil på figurene 31A og 31D ville øke, siden det trolig var flere som unnlot å ta stilling til disse utklippene og av den grunn fikk riktig svar.

Siden det er betydelig lettere å treffe riktig på de enkelte figurene enn spørsmålet i sin helhet, vil den gjennomsnittlige z-skåren være lavere og p-verdien høyere for hver av figurene. En korrelasjonsmatrise for riktig svar mellom figurene viser at de korrelerer signifikant på 0,1 % nivå med hverandre.

En krysstabell mellom besvarelsene for de enkelte figurene viser med unntak av B og C, en klar sammenheng. Ser vi derimot på B og C som begge gir en terning etter reglene, får vi følgende:

**Krysstabell mellom figurene B og C**

Count

| | | Figur C | | |
|---|---|---|---|---|
| | | 0 | 1 | Total |
| Figur B | 0 | 13 | 21 | 34 |
| | 1 | 25 | 56 | 81 |
| Total | | 38 | 77 | 115 |

Av de 115 som svarte på spørsmålet var det bare 56 som svarte riktig på begge og 46 som svarte riktig på den ene. Overraskende viser en Kji-kvadratverdi på 0,59 at det er liten sammenheng mellom besvarelsene for disse to figurene.

De tre første spørsmålene i oppgaven er alle problemstillinger rundt konstruksjon av en terning til spill. En krysstabell for spørsmålene vil vise hvordan besvarelsene på de respektive spørsmålene forholder seg til hverandre.

**Krysstabell mellom spørsmål 29, spørsmål 30 og spørsmål 31**

Count

| | | | spml.29 | | | | |
|---|---|---|---|---|---|---|---|
| spørs31 | | | 10 | 70 | 79 | 99 | Total |
| 0 | spml.30 | 10 | 21 | 2 | 0 | 0 | 23 |
| | | 73 | 1 | 5 | 1 | 0 | 7 |
| | | 74 | 3 | 2 | 0 | 0 | 5 |
| | | 79 | 4 | 3 | 3 | 0 | 10 |
| | | 99 | 10 | 7 | 0 | 1 | 18 |
| | Total | | 39 | 19 | 4 | 1 | 63 |
| 1 | spml.30 | 10 | 33 | 0 | | | 33 |
| | | 70 | 1 | 0 | | | 1 |
| | | 71 | 2 | 0 | | | 2 |
| | | 72 | 4 | 0 | | | 4 |
| | | 73 | 4 | 0 | | | 4 |
| | | 79 | 1 | 0 | | | 1 |
| | | 99 | 5 | 2 | | | 7 |
| | Total | | 50 | 2 | | | 52 |
| 99 | spml.30 | 10 | 3 | 0 | 1 | 0 | 4 |
| | | 73 | 1 | 0 | 0 | 0 | 1 |
| | | 79 | 0 | 1 | 1 | 0 | 2 |
| | | 99 | 4 | 1 | 1 | 15 | 21 |
| | Total | | 8 | 2 | 3 | 15 | 28 |

Krysstabellen viser at 33 elever, eller 23,1 %, svarte riktig på alle tre spørsmålene. Av disse var det 3 elever som hadde negativ z-skåre på hele testen. Dette viser igjen at enkelte elever som generelt gjør det dårlig på en tradisjonell matematikktest, kan føle seg mer hjemme på andre områder av matematikken. Det var overraskende å se at 17 elever som svarte riktig på både spørsmål 29 og spørsmål 31, svarte feil eller blankt på spørsmål 30. Spørsmål 31 hadde større vanskegrad. I tillegg til regelen "*summen av antall prikker på to motstående sider alltid er syv*" ved konstruksjon av terninger, måtte elevene også kunne se for seg hvilke sider som står i forhold til hverandre. Jeg antar at flere av de som ikke mestret spørsmål 30, men svarte riktig på spørsmål 31, ikke fikk med seg hvor enkelt problemet kunne løses. Tabellen viser også at alle de som svarte riktig på både spørsmål 30 og 31, også svarte riktig på spørsmål 29.

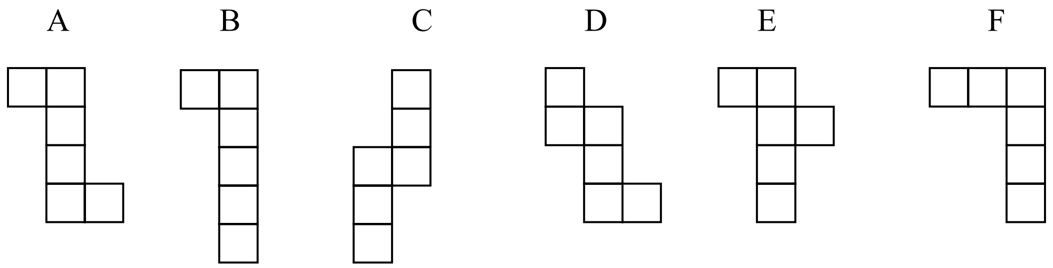
Vi ser også av tabellen at de 7 elevene som ble kodet 70,71 og 72 under spørsmål 30 viste at de mestret regelen for konstruksjon av terninger, men at de enten forvekslet vannrett og loddrett eller var usikre på hvilke sider som skulle regnes med

---

**Spørsmål 32:** TERNINGER

Figurene over er ett eksempel på hvordan et pappstykke kan klippes til for så å brettes til en terning. Det er flere andre måter pappstykket kan klippes til på.

Sett en ring rundt de av figurene nedenfor som du tror kan bli en terning:




---

Mange elever har i oppveksten, enten i skolesammenheng eller i fritiden, syslet med å klippe og lime sammen et pappstykke med det formål å få en terning, som er ett av de fem Platonske legemer (et polyeder der alle sideflatene er regulære mangekanter av samme type). De vil da også ha forskjellige erfaringer fra en slik syssel. Valg av figurer til dette spørsmålet sto mye i forhold til hva jeg forventet elevene ville mestre. Figurene A, C, D og E vil alle kunne klippes og brettes til en terning. To av disse, C og D antok jeg at ville være en større utfordring for de fleste. De to figurene som ikke kunne bli en terning, B og F, ble valgt med den begrunnelse at elevene forhåpentlig klart ville se at disse ikke kunne bli terning.

Dette spørsmålet hadde, i likhet med det foregående, den svakheten at det ikke ble etterfulgt av en tabell der elevene skulle avmerke sitt svar.

## Koder til spørsmål 32

| Koder | Frekvens | Prosent | Prosent av de som svarte | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 20 | 24 | 16,8 | 19,5 | 0,970 | 0,44 |
| 10 | 36 | 25,2 | 29,3 | 0,392 | 0,23 |
| | **Ikke riktig svar** | | | | |
| 70 | 30 | 21 | 24,4 | -0,059 | -0,03 |
| 71 | 24 | 16,8 | 19,5 | -0,489 | -0,22 |
| 72 | 7 | 4.9 | 5,7 | -0,294 | -0,077 |
| 73 | 2 | 1,4 | 1,6 | -0,667 | -0,08 |
| | **Ikke svart** | | | | |
| 99 | 20 | 14 | | -1,027 | -0,42 |

Her har jeg brukt kode 20 for de som har riktig på alle figurene og kode 10 for de som har riktig på fem av figurene. De andre kodene var for besvarelser som ikke ga poeng selv om de hadde én eller flere riktige. Kode 70 er de besvarelsene som har fire riktige, med tilsvarende tre riktige under kode 71, to riktige under kode 72 og én riktig under kode 73. Ingen av besvarelsene hadde null riktige selv om det teoretisk vil være ca. 1,5 % sjanse for det ved ren tipping.

Spørsmålet oppnådde en p-verdi på bare 0,29 selv om nesten 49 % av de som besvarte spørsmålet oppnådde poeng. Tabellen viser at både de under kodene 20 og 10 har en klar positiv gjennomsnittlig z-skåre og ”point-biserial”. Slår man sammen disse kodene vil ”point-biserial” bli på hele 0,53, høyere enn for hver av kodene siden korrelasjonen da også baseres på flere besvarelser. En korrelasjonsmatrise viser også her positiv korrelasjon for elevbesvarelsene mellom hver figur.

På spørsmål 31 var det en elev under **Ikke svart** som hadde positiv z-skåre, men på dette spørsmålet har alle 20 under **Ikke svart** negativ z-skåre. Vi ser også at antallet som oppga et svar, økte med 8. Tabellen viser også at de syv besvarelsene under kode 72, de som hadde to riktige figurer, oppnådde bedre gjennomsnittlig z-skåre enn de under kode 71. Årsaken var at tre av besvarelsene hadde en relativt høy z-skåre på hele testen.

Ser vi på besvarelsene for de enkelte figurene, får vi følgende tabell. Her er ikke de 20 som ikke besvarte spørsmålet tatt med. Dette var svake elever med en gjennomsnittlig z-skåre på –1,027.

## Koder til de respektive figurer i spørsmål 32

Kode **Ja** står her for kan bli en terning og kode **Nei** står for kan ikke bli en terning. Riktig svar er merket med *. Igjen har spørsmålsformuleringen den svakhet at **Nei** er det samme som at eleven ikke har satt ring rundt alternativet, og er derved enten et uttrykk for en vurdering av at alternativet ikke kan bli terning, eller at eleven ikke har tatt stilling til dette.

| **Koder til de respektive figurer** | **Frekvens** | **Prosent av de som svarte** | **Z-skåre** | **Point-biserial** |
|---|---|---|---|---|
| **Figur A** | | | | |
| Ja* | 98 | 79,7 | 0,334 | 0,49 |
| Nei | 25 | 20,3 | -0,486 | -0,23 |
| **Figur B** | | | | |
| Ja | 6 | 4,9 | -0,626 | -0,13 |
| Nei* | 117 | 95,1 | 0,208 | 0,44 |
| **Figur C** | | | | |
| Ja* | 53 | 43,1 | 0,434 | 0,34 |
| Nei | 70 | 56,9 | -0,036 | -0,04 |
| **Figur D** | | | | |
| Ja* | 65 | 52,8 | 0,472 | 0,43 |
| Nei | 58 | 47,2 | -0,174 | -0,15 |
| **Figur E** | | | | |
| Ja* | 91 | 74 | 0,355 | 0,47 |
| Nei | 32 | 26 | -0,366 | -0,20 |
| **Figur F** | | | | |
| Ja | 15 | 12,2 | 0,052 | 0,02 |
| Nei* | 108 | 87,8 | 0,183 | 0,32 |

I en sammensatt flervalgsoppgave der svaralternativene enten er **Ja** eller **Nei** er det naturlig å anta at det riktige svaret får størst oppslutning. Ser man på spørsmål 31 som også var en sammensatt flervalgsoppgave, ble tilslutningen til det riktige svaret betraktelig høyere for alle fire figurene. Tilsvarende ser man også på spørsmål 8.

Tabellen for de enkelte figurene på spørsmål 32 viser at figur D og spesielt figur C, skapte usikkerhet. Av de elevene som besvarte spørsmålet, var det bare 43,1 % som så at figur C ville ble en terning. Denne figuren var det for øvrig en av mine elever på Bredtveit fengsel som gjorde meg oppmerksom på. Den gjennomsnittlige z-skåren viser også at det var vesentlig flinke elever som oppga riktig alternativ. En frekvenstabell på de som valgte feil alternativ, viser at mange flinke elever bommet på figuren. Av de 70 som bommet, var det 29 som hadde positiv z-skåre på hele testen, av disse var det 10 som hadde over 1.

---

**Spørsmål 33:** TERNINGER

I spørsmål 31 og 32 så vi eksempler på hvordan pappstykker kan klippes til for å kunne brettes til en terning.

Har du selv et forslag til hvordan du kan klippe til et pappstykke slik at det kan brettes til en terning, i tillegg til de som er presentert i spørsmål 31 og 32?

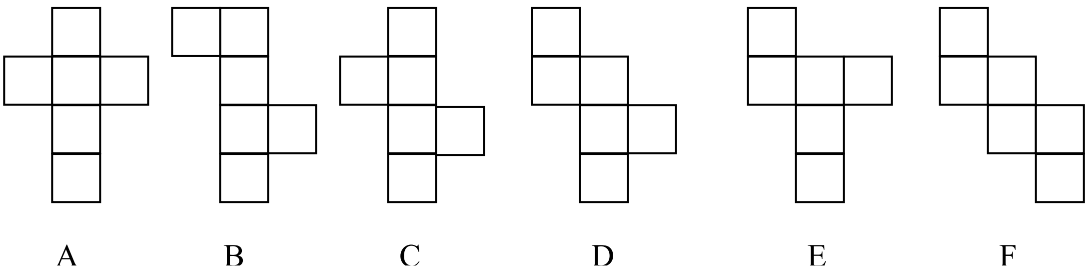
Tegn skisse her:

---

For å besvare dette spørsmålet bør elevene ha arbeidet med:

- rotasjon og speiling av figurer i planet
- å klassifisere figurer etter egenskaper og undersøke mønstre

Her er de siste seks alternativene:




Det er altså ikke mindre enn 11 forskjellige måter å klippe ut og lime sammen et pappstykke for å få en terning. Under spørsmålene 31 og 32 ble fem av disse presentert. I spørsmål 33 skulle elevene selv presentere et nytt. Jeg ser i ettertid at spørsmålet er upresist. Noen elever kan muligens ha forstått spørsmålet dit hen at et speilbilde eller en rotasjon av de tidligere figurene var nye forslag. Dette var ikke intensjonen ved spørsmålet. Jeg har derfor valgt å kode slike forslag under **Ikke riktig**. Svaralternativene henviser til de seks siste mulighetene som er vist over. Ingen av elevene presenterte alternativ F.

## Koder til spørsmål 33

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | Fig. A | 60 | 42 | 0,347 | 0,30 |
| 11 | Fig. B | 7 | 4,9 | 1,582 | 0,36 |
| 12 | Fig. C | 5 | 3,5 | 0,375 | 0,07 |
| 13 | Fig. D | 1 | 0,7 | 0,990 | 0,08 |
| 14 | Fig. E | 1 | 0,7 | 0,280 | 0,02 |
| | **Ikke riktig svar** | | | | |
| 70 | Tidligere presentert | 6 | 4,2 | -0,035 | -0,01 |
| 79 | Andre svar (umulig) | 6 | 4,2 | -0,449 | -0,09 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 57 | 39,9 | -0,564 | -0,46 |

Over 60 % av elevene besvarte spørsmålet og presenterte minst ett forslag på hvordan de mente et pappstykke kunne klippes og brettes sammen til en terning. Av disse var det 86 % som tegnet en skisse på et nytt forslag som kunne brettes til terning. 14 % tegnet skisser som enten var presentert tidligere eller ikke kunne bli en terning. Det var interessant å se at elevene samlet fant frem til fem av de seks gjenstående mulighetene. Spørsmålets p-verdi ble da på 0,52.

Tabellen viser at det var positiv gjennomsnittlig z-skåre på alle de riktige forslagene. Dette gjelder spesielt de under kode 11 der alle besvarelsene hadde positiv z-skåre på hele testen. Tabellen viser også at de to som ble kodet henholdsvis 13 og 14, begge hadde positiv z-skåre på hele testen.

En frekvenstabell for z-skårene for både kode 10 og kode 12 viser at også flere svake elever kom med riktige forslag. Det var i tillegg flere flinke elever som ikke gjorde det enten ved at de ikke besvarte spørsmålet eller de kom med forslag som var en rotert eller speilvendt tegning av de tidligere fem figurene. På tross av dette, oppnådde de som oppga godkjent figur, gjennomsnittlig z-skåre på 0,474 og ”point-biserial” på 0,49. Tatt i betraktning den ideelle p-verdien, diskriminerte spørsmålet bra. Sammenlignet med for eksempel spørsmål 21 som hadde omtrent samme vanskegrad og ble besvart riktig av stort sett de flinkere elevene, var både den gjennomsnittlige z-skåren og ”point-biserial” vesentlig høyere på det spørsmålet enn på dette.

Under kode 70 var det seks forslag som alle var en speilvendt eller rotert skisse av de fem tidligere presenterte figurer. To av disse var fra elever med høy skåre på hele testen. Det førte til at den gjennomsnittlige z-skåren for denne gruppa ble like under null. De forslagene som ble kodet 79 var skisser som umulig kunne brettes sammen til en terning.

De 57 elevene som ikke besvarte spørsmålet og derfor ble kodet 99, hadde som ventet negativ gjennomsnittlig z-skåre, –0,564. Men i denne gruppa var det også syv elever som skåret bra på hele testen.

Spørsmålene 32 og 33 er beslektet. På begge spørsmålene skal elevene vise hvordan de enkelte sidene på figurene forholder seg til hverandre når de brettes sammen til terning.

**Krysstabell mellom spørsmål 32 og spørsmål 33**

Count

| | | Spørsmål 32 | | | | Total |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 99 | |
| Spørsmål 33 | 10 | 29 | 19 | 12 | | 60 |
| | 11 | | 2 | 5 | | 7 |
| | 12 | | 3 | 2 | | 5 |
| | 13 | | 1 | | | 1 |
| | 14 | | 1 | | | 1 |
| | 70 | 5 | 1 | | | 6 |
| | 79 | 5 | | 1 | | 6 |
| | 99 | 24 | 9 | 4 | 20 | 57 |
| Total | | 63 | 36 | 24 | 20 | 143 |

Det som slår en umiddelbert når en studerer krysstabellen, er at det ikke var noen av de 20 elevene som ikke besvarte spørsmål 32, som forsøkte seg på eller kom med forslag på nye terningkonstruksjoner på spørsmål 33 (19 av disse besvarte heller ikke spørsmål 31). Tvert imot fikk de følge av 37 nye elever. Hva kan forklaringen være på at så mange ikke kom med et forslag? Det er mulig at flere elever oppfattet spørsmål 33 "vanskeligere" selv om vanskegraden gitt ved p-verdien gir 0,52 på spørsmål 33 mot 0,29 for spørsmål 32. Årsaken kan også ligge i tidsnød der noen elever ikke fikk tid til å tenke ut nye figurer. Dette styrkes ved at antallet under **Ikke svart** økte med 43 i forhold til forrige spørsmål. Av disse var det 11 med positiv z-skåre. Flere av disse skåret også høyt. Dersom spørsmålet var gitt tidligere i pilottesten, ville kanskje flere elever kommet med forslag på figurer.

En annen forklaring kan være at det for flere elever er en større utfordring å komme med konkrete forslag som i dette tilfelle å presentere en ny figur, enn å besvare et spørsmål uavhengig av oppgaveformatet. Denne forklaringen blir støttet av at ingen andre spørsmål i pilottesten oppnådde så høy p-verdi tatt i betraktning av den lave svarprosenten.

Krysstabellen mellom spørsmålene viser en klar sammenheng i svarmønsteret, noe også en Kji-kvadrattest for de som svarte riktig og de som ikke svarte riktig, viser med verdien 8,13. Men tabellen viser også at det var fire elever som besvarte spørsmål 32 helt riktig, men ikke besvarte spørsmål 33. Tre av disse hadde negativ z-skåre på hele testen, mens den siste hadde en meget høy skåre. Det er da pussig at de ikke selv kom med forslag på figur når de samtidig så hvilke av figurene i spørsmål 32 som kunne bli til en terning. Dette forsterkes når man samtidig ser at 29 elever som ikke oppnådde poeng på spørsmål 32, foreslo en figur som bare var ”ubetydelig” forskjellig fra utgangsfiguren i spørsmål 31.

Tabellen viser også at de fire andre alternativene på figurer foruten figur A, alle ble presentert av elever som enten oppnådde 1 eller 2 poeng på spørsmål 32.

### 7.7.1 Oppsummering rundt oppgaven Terninger

Som ventet for en oppgave med kontekst nær elevene, var svarprosenten høy. Selv om oppgaven ble presentert sent i testen, var det bare 12 elever som ikke svarte på noen av spørsmålene. Dette viser igjen at på oppgaver som står elevene ”nær”, øker svarprosenten. 84 % av elevene forsøkte seg på noen eller alle spørsmålene i denne oppgaven. Det var til sammen 20 elever som svarte riktig på alle spørsmålene inkludert seks av elevene som ”bare” hadde fem riktige på spørsmål 32. De hadde alle positiv z-skåre på hele testen, den gjennomsnittlige for disse elevene ble på 1,245. Oppgaven sett under ett fikk ”point-biserial” på 0,50, som er høyt tatt i betraktning at den er basert på bare 20 elever.

Av mine forslag var det kanskje denne oppgaven jeg likte best. Spørsmålene ga utfordringer som elevene i mindre grad møter på tradisjonelle tester og undersøkelser under faget matematikk. Dette gjelder spesielt spørsmål 33 der elevene selv skal komme med forslag til hvordan et pappstykke kan klippes til for å kunne brettes til en terning. Oppgaver der elevene selv skal komme med forslag, har fra og med år 2000 kommet med i avgangsprøven i matematikk. En gjennomgang av samtlige oppgaver som ble gitt til Generalprøven, viser at ingen av disse matematikkoppgavene forutsetter at elevene selv skal komme med forslag. Én årsak kan være at det er mer krevende å rette slike oppgaver. Ved retting av oppgaver til PISA er sensorreliabiliteten viktig. Jeg tror allikevel at PISA-undersøkelsen ville vært tjent med å ha med en oppgave av denne type, slik at elevene fikk anledning til å vise sin ”kreativitet” ved selv å komme med forslag i motsetning til de oppgaver som i hovedsak fordrer entydige svar. Det er også mulig å utvikle sikre retteprosedyrer for slike oppgaver ved at det på forhånd er satt opp godkjente svaralternativer.

Det var også gledelig å se at flere elever som ikke besvarte enklere spørsmål i oppgaven tok utfordringen med de to siste spørsmålene. Av de 16 elevene som ikke besvarte første spørsmål i oppgaven, spørsmål 29, var det fire som besvarte spørsmål 32. To av disse svarte riktig på fem av figurene. To kom også med forslag til andre måter å brette et pappstykke til en terning på i spørsmål 33. Tilsvarende var det 9 elever som svarte feil på spørsmål 29 som oppga riktig forslag på spørsmål 33. Dette var i hovedsak meget svake elever med gjennomsnittlig z-skåre på –0,626. En forklaring på dette kan være at flere svake elever blir ”tent” på oppgaver som ligger til siden for de tradisjonelle oppgavene de til daglig møter i skolen.

Både spørsmålene 30 og 31 kom med på Generalprøven, men spørsmål 31 gjennomgikk en forenkling ved at det ble påført prikker på de sidene som ble vist blanke i pilottesten. I tillegg ble spørsmål 31 klart forbedret ved at den ble etterfulgt av en tabell der elevene skulle markere hvorvidt utklippet kan foldes sammen til en terning eller ikke.

Ved utvikling av denne oppgaven tok jeg utgangspunkt i de felles mål for matematikkfaget som fremkommer i læreplanverket (L97). Flere av spørsmålene stimulerer elevene til å bruke fantasi og kunnskaper for å finne løsningsalternativer. Dette legger jeg til grunn for oppgavens innholdsvaliditet i forhold til L97. Jeg anser også at flere av spørsmålene har denne validiteten i forhold til PISA. At flere av spørsmålene ble benyttet i Generalprøven, styrker dette.

# 8. Reliabilitetsbetraktninger

Det henvises til p.4.2 og p.4.2.1 der det er redegjort for reliabilitet og den teknikk som kan benyttes for å måle intern konsistens for spørsmålene i denne pilottesten. Tabellen nedenfor er basert på at riktig svar, kode 10 er gitt 1 poeng, eventuelt 2 poeng der kode 20 også er benyttet. Poengene er summert for alle spørsmål. Deretter er følgende tabell regnet ut:

**Item-Total Statistics**

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| q1 | 15,67 | 70,757 | ,196 | ,913 |
| q2 | 15,66 | 70,816 | ,197 | ,913 |
| q3 | 15,97 | 68,471 | ,353 | ,912 |
| q4 | 16,06 | 68,073 | ,383 | ,911 |
| q5 | 15,99 | 68,866 | ,295 | ,913 |
| q6 | 16,11 | 67,565 | ,442 | ,911 |
| q7 | 16,12 | 67,063 | ,505 | ,910 |
| q8 | 16,38 | 68,039 | ,450 | ,911 |
| q9 | 16,50 | 69,266 | ,376 | ,911 |
| q10 | 16,59 | 70,272 | ,340 | ,912 |
| q11 | 16,23 | 67,939 | ,405 | ,911 |
| q12 | 16,35 | 68,116 | ,422 | ,911 |
| q13 | 16,06 | 63,651 | ,637 | ,908 |
| q14 | 16,43 | 68,120 | ,308 | ,913 |
| q15 | 15,94 | 68,355 | ,377 | ,911 |
| q16 | 15,74 | 69,855 | ,285 | ,912 |
| q17 | 16,41 | 67,103 | ,618 | ,909 |
| q18 | 16,24 | 67,774 | ,427 | ,911 |
| q19 | 16,02 | 68,232 | ,369 | ,912 |
| q20 | 16,59 | 70,623 | ,259 | ,913 |
| q21 | 16,13 | 65,562 | ,695 | ,907 |
| q22 | 16,15 | 65,399 | ,717 | ,907 |
| q23 | 16,15 | 65,535 | ,699 | ,907 |
| q24 | 16,15 | 65,915 | ,650 | ,908 |
| q25 | 15,87 | 70,144 | ,158 | ,914 |
| q26 | 16,27 | 68,312 | ,367 | ,912 |
| q27 | 16,51 | 69,195 | ,400 | ,911 |
| q28 | 16,41 | 69,736 | ,222 | ,913 |
| q29 | 15,95 | 66,948 | ,560 | ,909 |
| q30 | 16,21 | 66,716 | ,556 | ,909 |
| q31 | 16,27 | 66,760 | ,566 | ,909 |
| q32 | 16,04 | 64,815 | ,494 | ,911 |
| q33 | 16,11 | 67,537 | ,446 | ,911 |
| q34 | 16,21 | 66,308 | ,608 | ,908 |
| q35 | 16,35 | 67,144 | ,557 | ,909 |
| q36 | 16,34 | 66,168 | ,684 | ,907 |
| q37 | 16,55 | 69,587 | ,404 | ,911 |
| q38 | 16,56 | 69,783 | ,377 | ,912 |

Som beskrevet i p.5.3, ble riktig svar gitt poeng. Ut fra disse poengene vil den første kolonnen til høyre for spørsmålet, angi gjennomsnittspoeng for testen dersom det respektive spørsmålet utelates. Neste kolonne angir variansen til disse poengene. Årsaken til den høye variansen er variasjonsbredden for antall poeng elevene oppnådde på testen (fra 1 til 38). Kolonnen *Corrected Item-Total correlation* angir tilsvarende korrelasjon mellom de enkelte spørsmål i forhold til summen av de andre spørsmålene som er med i testen. Verdien i denne kolonnen ligger nær opp til "point-biserial" korrelasjon for riktige svar på de respektive spørsmål. Det er en korrelasjon der den ene variabelen er dikotom og den andre på intervall- eller forholdstallnivå. For å regne ut denne koeffisienten kan vi benytte samme formel som ved utregning av "point-biserial" (se p.4.1.4). Vi må først korrigere gjennomsnitt og standardavvik for hele testen ved å fjerne det bidrag det respektive spørsmål gir. Den siste kolonnen angir Cronbachs alfa for hele testen dersom det respektive spørsmålet ble utelatt.

Vi ser av kolonnen at flere spørsmål korrelerer svakt med de andre spørsmålene i pilottesten. Dette gjelder spesielt de to første spørsmålene som fikk en svært høy p-verdi der omtrent alle, også de som gjorde det dårlig på testen, svarte riktig.

Vi ser også at spørsmål 5 hadde en svak korrelasjon. Av de 91 som valgte riktig alternativ på denne flervalgsoppgaven, var det 43 som hadde negativ z-skåre på hele testen. Samtidig var det 14 med positiv z-skåre som ikke valgte dette alternativet.

Spørsmål 14, som var fra kompetanseklasse 3, fikk en lav p-verdi på bare 0,1. Årsaken til den svake korrelasjon kan, ved siden av den lave p-verdien, være at tre elever med negativ z-skåre svarte helt riktig på spørsmålet og ble da kodet 20. Det var også mange elever som gjorde det bra på testen som helhet, men som svarte galt her og ble kodet under **Ikke riktig svar**.

Den svake korrelasjonen på spørsmål 20 skyldes i hovedsak at bare fem elever svarte riktig på spørsmålet. Ser vi samtidig på spørsmål 10 der seks elever svarte riktig, fikk dette spørsmålet noe høyere korrelasjon for riktig svar. Her ser vi at antallet som svarer riktig på et spørsmål, ikke minst når det er få som i dette tilfellet, har stor betydning for korrelasjonskoeffisienten.

Selv om spørsmål 25 fikk en p-verdi på 0,76, korrelerte den svakest av alle spørsmål i testen. Grunnen til det kan vi se fra en frekvenstabell for z-skårene for de som svarte riktig. Så mange som 58 av de 109 som svarte riktig (53,2 %), hadde negativ z-skåre på hele testen. Samtidig var det 10 av de 22 som ble kodet under **Ikke riktig svar** som hadde positiv z-skår. En mulig årsak til denne skjevheten er nærmere omtalt i diskusjonen rundt oppgaven *Kast med to terninger.*

Vi ser også at spørsmål 28 korrelerte svakt med de andre spørsmålene i testen. Dette er en flervalgsoppgave der den ene distraktoren fikk størst tilslutning. 1/3 av de som valgte riktig alternativ hadde negativ z-skåre på testen. Samtidig valgte 36 elever med positiv z-skåre en av distraktorene. De to andre spørsmålene i oppgaven *Kast med to terninger*, hadde også svak korrelasjon med hele testen, om ikke så svak som spørsmål 25 og spørsmål 28.

For å bestemme reliabiliteten kan det være ønskelig gjennomføre testen to ganger. Korrelasjonen mellom elevbesvarelsene på disse to testene ville gitt oss en reliabilitets-koeffisient. Å gjennomføre to slike tester kan ofte ha praktiske ulemper som for eksempel at det er mer arbeidskrevende. Dessuten er det fare for at elevene vil huske sine gamle svar dersom den andre testen var en gjentagelse av den første.

Denne pilottesten baserte seg på at elevene fikk én test. Det er da også flere andre metoder å regne ut reliabilitetskoeffisienten på. En mye brukt metode er *Split-half* der spørsmålene i pilottesten deles i to og man regner ut korrelasjonen mellom disse to halvdeler.

> *Other internal-consistency measures of reliability do not require splitting the test into halves and scoring each half separately. These procedures assess the inter-item consistency, or homogeneity, of items. (Ary 1996, s.283)*

Hva ble så Cronbachs alfa (se p.4.2) for denne pilottesten?

**Reliability Statistics**

| Cronbach's Alpha | N of Items |
|---|---|
| ,913 | 38 |

Vi ser av utskriften fra SPSS at den ble høy, spesielt tatt i betraktning at også flere av besvarelsene korrelerte svakt med de andre besvarelsene i testen. Dette viser i ettertid at flere av spørsmålene kunne vært utelatt uten å påvirke Cronbachs alfa i nevneverdig grad. Dette forsterkes av at noen elevers motivasjon antagelig ble redusert på slutten av testen.

Dersom man fjerner de 8 spørsmålene (se kommentar over) som korrelerte svakest med de andre spørsmålene, ville alfa øke til 0,917. Tilsvarende ville alfa reduseres til 0,854 dersom de 8 spørsmålene som korrelerte høyest, ble fjernet. Blant disse spørsmålene er de fire fra oppgaven *Fibonaccitallene*. En korrelasjonstabell viser klart at alle spørsmålene korrelerer svært høyt med hverandre i den oppgaven. Dette viser i hovedtrekk at de er for "like".

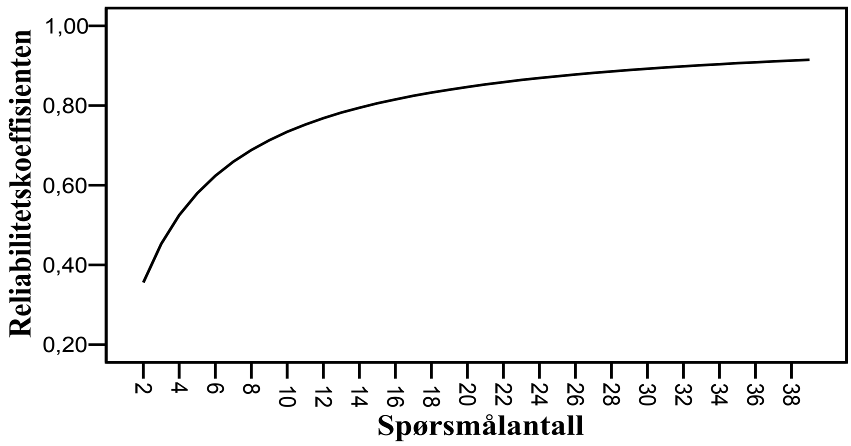
Spearman og Brown presenterte en formel for hvordan reliabiliteten teoretisk vil øke dersom en test forlenges et visst antall ganger. Det er rimelig å anta at tilsvarende også vil gjelde ved reduksjon av antall spørsmål i testen. Tar man utgangspunkt i reliabilitetskoeffisienten på 0,913 og antallet 38 spørsmål i pilottesten, vil en omarbeidet utgave av Spearman-Browns formel se slik ut:

$$p_{xx'} = \frac{\frac{a}{38}0{,}913}{1+(\frac{a}{38}-1)0{,}913}$$

der $p_{xx'} =$ testens reliabilitet

$a =$ spørsmålantallet

**Reliabilitetskoeffisienten ved redusert antall spørsmål**




Vi ser av grafen at reliabilitetskoeffisienten øker som en funksjon av antall spørsmål i testen. Økningen avtar med stigende antall spørsmål og har tilnærmet flatet ut når vi nærmer oss de 38 spørsmålene i denne pilottesten. Tilsvarende vil en moderat reduksjon av antall spørsmål fra de opprinnelige38, ha liten innvirkning på reliabilitetskoeffisienten. Selv en halvering av antallet vil gi en verdi for $\alpha$ på 0,84.

En korrelasjonsmatrise for alle 38 spørsmålene gir 703 kombinasjoner. Av disse er det 688 med positiv korrelasjon og bare 15 med negativ. Dette er en medvirkende årsak til den relativt høye Cronbachs alfa.

Et annet spørsmål ved reliabiliteten er sensureringen. Sensorreliabiliteten er et mål for hvor pålitelig retting og koding av svarene er gjennomført, og vil da være et uttrykk for samsvaret mellom flere sensorers retting av spørsmålene. Sensorreliabiliteten er større ved flervalgsoppgaver og kortsvaroppgaver enn ved åpne oppgaver der eleven skal vise utregninger eller begrunne sitt svar. Det var få åpne oppgaver i pilottesten, og i tillegg kodet jeg oppgavene to ganger. Jeg anser derfor at denne reliabiliteten er høy.

# 9. Konklusjon

Utgangspunktet for denne hovedoppgaven var oppgaveutvikling til PISA 2003. Både analysen av oppgavene og responsen fra PISA, viste at ikke alle oppgavene var like velegnet til dette formålet. Det var også av interesse å se hvorvidt oppgavene også hadde andre anvendelige sider. Det var én grunn til at analysedelen ble en forholdsvis omfattende del av hovedoppgaven.

Oppsummeringen etter hver oppgave i analysedelen viste klare svakheter både på enkelte spørsmål og på de respektive oppgavers oppbygning. Dette burde vært bedre gjennomarbeidet til pilottesten. En av testens intensjoner var å registrere slike svakheter som også la grunnlaget til forbedringer både på spørsmålene og den innledende teksten til hver av oppgavene. Dette ble gjort på de seks oppgavene som ble sendt til PISA.

Oppgaveforslagene skulle utvikles etter retningslinjer fra *Mathematics item development for PISA 2003* og det var derfor naturlig å sette en av hovedproblemstillingen rundt **Fungerer mine oppgaveforslag bra psykometrisk og oppfyller de kravene til rammeverket til PISA**?

Oppgavene i sin helhet dekket i hovedsak de innholdsmessige sidene ved de sentrale ideer i rammeverket. Dette bekreftes ved at flere av oppgavene kom med i Generalprøven. Jeg fikk også med oppgaver til PISA-undersøkelsen i 2003. De kvantitative målemetoder jeg benyttet i analysedelen viste også at flere av oppgavene er gode ”konstruksjoner” og er anvendelig også i andre testsammenheng. Jeg anser derfor at analysen av elevbesvarelsene bekrefter at mine oppgaveforslag fungerer bra psykometrisk.

Jeg hadde også andre intensjoner med oppgavene. Jeg ønsket å kunne bruke reviderte utgaver av disse oppgavene senere i min egen undervisning, enten som diagnostiske oppgaver eller for å evaluere elevenes kunnskaper og ferdigheter. Min andre hovedproblemstilling var om **mine oppgaveforslag lå innenfor L97**. Dette ble for flere av spørsmålene bekreftet ved avgangsprøven som ble gitt noen uker etter pilottesten. Vi så i analysedelen at flere av spørsmålene lå nær opp til oppgaver gitt til denne og etterfølgende eksamener.

Flere av oppgaveforslagene har jeg allerede benyttet i min egen undervisning enten for å introdusere et nytt tema eller for å kartlegge elevenes kunnskaper innen begrensede deler av matematikkfaget. På den måten har jeg også undersøkt oppgavenes kriterievaliditet.

Ikke alle oppgavene viste seg å være velegnet til den undersøkelsen PISA skulle gjennomføre. Var det andre viktige sider ved slike oppgaver? Et viktig funn ved analysen var hvorvidt noen av oppgavene kunne ha en diagnostisk funksjon. Derfor var en av problemstillingene **om noen av mine forslag fungerer som en diagnostisk test**.

Analysen av besvarelsene i oppgaven *Kast med to terninger* viser at den var lite egnet til PISA-undersøkelsen eller en prøve for å rangere elevene. Det var bare seks elever som svarte riktig på alle fire spørsmålene, selv om 74,8 % av elevene besvarte alle. Den store andelen under **Ikke riktig svar** kan til en viss grad skyldes at flere elever ikke leste innledningen til oppgaven grundig nok. En annen viktig årsak var elevenes misoppfatninger rundt begrepet sannsynlighetsregning. Det er vesensforskjell på feil elevene gjør på et spørsmål og de misoppfatninger de har. Ved å følge svarmønstret mellom spørsmålene ved hjelp av krysstabeller, var det lett å identifisere elevenes misoppfatninger rundt sannsynlighetsbegrepene. I oppsummeringen etter oppgaven begrunnet jeg hvorfor særlig denne ville fungere som en diagnostisk oppgave.

> *Sentralt i begrepet matematikk i PISA står det å kunne anvende matematikken i ulike situasjoner og kontekster… Noen av situasjonene kan være "nært" knyttet til elevenes erfaringer, andre mer fjernt. (Lie, Kjærnsli, Roe & Turmo 2001, s.59).*

Min tredje problemstilling var å registrere om det **er tendens til at oppgaver med kontekst "nær" eleven, øker svarprosenten.**

Vi ser av analysen at ved oppgaver med kontekst "fjernt" fra eleven ble svarprosenten redusert betraktelig i forhold til oppgaver med kontekst hentet fra *Personlig liv*. I den første oppgaven, *Poteter*, var det ingen elever som ikke svarte på minst ett av spørsmålene. Vi så også at i de to andre oppgavene med kontekst "nær" eleven, *Kast med to terninger* og *Terninger*, var det bare henholdsvis 9 og 12 elever som ikke besvarte noen av spørsmålene selv om oppgavene ble presentert på slutten av testen. Ser man så på de oppgaver med kontekst "fjernt" fra eleven, har antallet som ikke svarer på noen av spørsmålene, økt betraktelig. Det var 30 elever som ikke svarte på noen av spørsmålene i oppgaven *Fibonaccitallene*. Videre var det så mange som 55 elever som ikke svarte på noen av spørsmålene i oppgaven *Totallsystemet* (se vedlegg).

Vi så også at oppgaven *Barnefødsler* fikk en høy svarprosent på de to første spørsmålene. Det var bare fire elever som ikke svarte på noen av spørsmålene i oppgaven. Samtidig var p-verdien på alle spørsmålene høyere på oppgaven *Fibonaccitallene* enn på spørsmålene i denne oppgaven. Dette kan tyde på at oppgaver hentet fra *Nærmiljø og samfunn* engasjerer flere elever til å oppgi et svar enn oppgaver som sto "fjernere" fra eleven. Tilsvarende ser vi at de to første spørsmålene i oppgaven *Fahrenheit – Celsius* også har høy svarprosent selv om den reduseres på de siste spørsmålene.

Som nevnt under problemstillingen (se p.1.1.1) kan det være flere årsaker til at en elev ikke besvarer enkelte av spørsmålene i en oppgave. Hvilken plassering spørsmålet har i en test vil ha en innvirkning ved for eksempel tidsnød. En oppgave med en omfattende innledning i tillegg til at den fremstår som "vanskelig" vil også påvirke svarprosenten. Dette tatt i betraktning, er min konklusjon at vi ser en klar tendens til at oppgaver med kontekst "nær" eleven øker svarprosenten.

Tidligere PISA-undersøkelser viser store forskjeller mellom jenter og gutters gjennomsnittsskåre i lesing generelt. Derfor tok jeg hensyn til dette ved utvikling av oppgaver og unngikk derfor på noen av oppgavene å benytte for mye tekst ved innledningen. En viktig side ved analysen av besvarelsene var å registrere om det likevel er eventuelle forskjeller i gutters og jenters prestasjoner som kan forklares med oppgavens formulering.

Vi så i analysen at det var signifikante forskjeller i besvarelsen mellom kjønnene i tre av spørsmålene. På både spørsmål 4 og spørsmål 13 gjorde jentene det bedre enn guttene. Tilsvarende samme forskjeller ble ikke registrert på de øvrige spørsmålene i de samme oppgavene. Siden de andre spørsmålene baserte seg på samme innledende tekst eller diagram, ble det også i analysen konkludert med at forskjellene ikke var grunnet i oppgaveformuleringen.

På spørsmål 27 gjorde guttene det signifikant bedre enn jentene. Spørsmålet fikk en lav p-verdi som kan være en av grunnene til denne forskjellen. Siden Kji-kvadratfordelingen bygger på normalfordelingen, vil påliteligheten til en test svekkes dersom cellene er for små. I tillegg viste også elevenes besvarelse en stor usikkerhet rundt de fleste spørsmålene i oppgaven. Selv om den innledende teksten kan ha vært noe av årsaken til denne usikkerheten, viser besvarelsene på de andre spørsmålene i oppgaven ingen forskjell mellom gutter og jenter. Jeg anser at oppgaven er lite egnet til å teste elevenes kunnskaper i sannsynlighetsregning og vil derfor legge liten vekt på denne forskjellen. Min konklusjon blir derved at det ikke er forskjellige prestasjoner mellom kjønnene grunnet i oppgaveformuleringen.

På slutten av oppgavesettet ble elevene bedt om å krysse av for den innsats de hadde lagt ned i testen. Det var 131 elever som krysset av på ”*Innsatstermometeret*". Det er innlysende at koeffisienten bør ligge vesentlig under den perfekte korrelasjonen på 1. Skulle den ligge nær 1 vil det gi et inntrykk av at elevprestasjonen sto mer i forhold til den oppgitte innsatsen enn deres kunnskaper og ferdigheter. Korrelasjonen mellom poengskåren og ”innsats” viser følgende:

**Korrelasjon mellom innsats og skåre**

| | | Innsats | skåre |
|---|---|---|---|
| Innsats | Pearson Correlation | 1 | ,565** |
| | Sig. (2-tailed) | | ,000 |
| | N | 131 | 131 |
| skåre | Pearson Correlation | ,565** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 131 | 143 |

**. Correlation is significant at the 0.01 level (2-tailed).

Som nevnt i p.6.3 vil det foreligge en usikkerhet ved en slik sammenlikning. Dette gjelder først og fremst hvorvidt elevene hadde forstått ”*Innsatstermometeret*” korrekt og ikke krysset av etter hvordan de selv følte de hadde prestert. Som vi ser av tabellen ble denne korrelasjonen 0,565 (signifikant på 0,01 nivå). Dette viser at elevenes poengskåre på testen korrelerer rimelig bra med den oppgitte innsats til de 131 elevene som krysset av.

# Referanser/Litteraturliste


Angell, C. (1996): Elevers fysikkforståelse. En studie basert på utvalgte fysikkoppgaver i TIMSS. Dr. scient-avhandling ved UIO, Norge

Angell, C. & Lie, S. (1993): Er eksamen rettferdig? Senter for lærerutdanning og Skoletjeneste, Universitetet i Oslo

Ary, D., Jacobs, L. & Razavieh, A. (1996): Introduction to Research in Education. 5. utg.. Harcourt Brace College Publishers

Brekke, G. (1995): Kartlegging av matematikkforståelse. Introduksjon til diagnostisk undervisning i matematikk .Telemarksforskning - Notodden

Brekke, G., Kobberstad, T. Lie, S. & Turmo, A. (1999): Hva i all verden kan elevene i matematikk? Oppgaver med resultater og kommentarer. Universitetsforlaget, Oslo

Cockcroft, W.H. (1982): Mathematics Counts-Report of the Committee of Enquiry into the Teaching of Mathematics in Schools. Department of Education and science, London

Erstad, G. & Bjørnsgård, I. (2000): Matematikk-nytt nr.6. Aschehoug, Oslo

Gjone, G. (1994): Matematikkundervisningen i etterkrigstidens enhetsskole – belysning av de kulturelle og demokratiske perspektiver. Foredrag holdt på Nordisk Forskersymposium

Kleve, B. (1994): En testteoretisk og diagnostisk analyse av flervalgsoppgaver matematikk fra TIMSS' pilottest. Hovedfagsoppgave, ILS, UIO, Oslo

KUF (1996): Læreplanverket for den 10-årige grunnskolen. KUF, Oslo

Lie, S., Kjærnsli, M., Roe, A. & Turmo, A. (2001): Godt rustet for framtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv. ILS, UIO, Oslo

Læringssenteret (2001a): Sensorveiledning. Skriftlig avgangsprøve i matematikk.

Læringssenteret (2001b): Avgangsprøven i matematikk for grunnskolen.

Niss, M. (1994): Mathematics in society. Kluwer Academic. Publishers, Dordrecht

Niss, M & Jensen, T.H. (2002): Kompetencer og matematiklæring. Ideer og inspiration til udvikling af matematikundervisning i Danmark. Undervisningsministeriet, Uddannelsesstyrelsens temahæfteserie nr. 18, København

Normalplanutvalget av 1967 (1970): Forarbeidet til normalplan for grunnskolen. Aschehoug, Oslo

OECD (2000): Measuring Student Knowledge and Skills. OECD PUBLICATIONS, Paris

OECD/PISA (2000): Mathematics item development for PISA 2003 and item submission guidelines. Max Planck Institute, Berlin

OECD/PISA (2001a): Draft mathematics framework for OECD/PISA 2003. Consultation Document

OECD/PISA (2001b): Draft framework for the PISA 2003 mathematics assessment. Nijmegen, Nederland

Olsen, R.V., Turmo, A. & Lie, S. (2001): Learning about students' knowledge and thinking in science through large-scale quantitative studies. European Journal of Psychology of Education 2001, Vol. XVI, nr 3

Ostad, S. (1993): Matematikk og matematikkvansker. Artikkelsamling, Hosle

Osterlind, S. (1992): Constructing test items. University of Missouri-Colombia

Sjøberg, S. (1998): Naturfag som allmenndannelse. Gyldendal, Oslo

Skaalvik, E.M & Skaalvik, S. (1996): Selvoppfatning, motivasjon og læringsmiljø, TANO

Tamir, P. (1990): Justifying the selection of answers in multiple choice items. International Journal of Science Education, 12, 5

Tietze, U-P. (1994): Mathematical curricula and the underlying goals. Kluwer Academic. Publishers, Dordrecht

Wedege, T. (2000): Matematikholdige kompetencer og kvalifikationer. Center for forskning i matematiklæring, Danmark

# Vedlegg

## TOTALLSYSTEMET

Kunnskap om tall og tallsystemer er viktig for å forstå hvordan en datamaskin fungerer. I dag er de aller fleste datamaskiner digitale, det vil si at de baserer seg på totallsystemet bygget opp av tallene 0 (ikke strøm) og 1 (strøm).

Fordelen med slike maskiner er at de ikke tolker signaler feil. Skulle maskinen bruke et titallsystem, som vi til daglig bruker, ville maskinen ha problemer med å sette grenser mellom alle ti tallene. En liten unøyaktighet i en av komponentene er nok til at maskinen tolker verdien feil.

Totallsystemet er i likhet med titallsystemet et posisjonssystem. I et titallsystem brukes de 10 sifrene fra 0 til 9. Fra tallet 10 brukes en kombinasjon av to sifre fram til tallet 99. Slik fortsetter vi med en kombinasjon av tre sifre, fire sifre osv. Vi kjenner alle disse tallene ganske godt fordi vi bruker dem hver dag, men det er ikke sikkert at vi alle tenker så ofte på hva hvert av tallene i f. eks. 3967 betyr. Leser vi tallene bakfra står det første sifferet for antallet enere (sju), det neste sifferet står for antallet tiere (seks), den neste plassen er hundrerplassen (ni) og tretallet står på den plassen som forteller hvor mange tusener vi har.

Vi konstruerer tall i totallsystemet på samme måte, men vi bruker bare de to sifrene 0 og 1. I dette tallsystemet er det da slik at det bakerste tallet står på enerplassen, det neste på toerplassen, deretter følger henholdsvis firer-plassen, åtter-plassen, seksten-plassen osv. Dette fører da til at tallene fort blir veldig lange. Tallet 5 i titallssystemet må vi f. eks. bruke tre sifre for å skrive i totallsystemet. Det skrives 101. Forklaringen er at fem består av én firer, ingen toere og én ener.

---

**Spørsmål 34:** TOTALLSYSTEMET

Tabellene viser en del av titallsystemet, og de tilsvarende tallene i totallsystemet. Sett inn de tallene i totallsystemet som mangler.

| Titallsystemet | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Totallsystemet | 0 | 1 | 10 | 11 | | 101 | | 111 | 1000 | |

## Koder til spørsmål 34

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 100..110…1001 | 60 | 42,0 | 0,756 | 0,65 |
| | **Ikke riktig svar** | | | | |
| 70 | Regnefeil | 10 | 7,0 | -0,205 | -0,06 |
| 79 | Andre svar | 12 | 8,4 | -0,587 | -0,18 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 61 | 42,7 | -0,594 | -0,51 |

---

**Spørsmål 35:** TOTALLSYSTEMET

Gjør om tallet 14 i titallsystemet til totallsystemet.

…………………………………………………………………………

---

## Koder til spørsmål 35

| Kode | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 1110 | 40 | 28,0 | 0,949 | 0,59 |
| | **Ikke riktig svar** | | | | |
| 70 | 0111 | 4 | 2,8 | 1,079 | 0,18 |
| 79 | Andre svar | 27 | 18,9 | -0,048 | -0,02 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 72 | 50,3 | -0,569 | -0,58 |

**Spørsmål 36:** TOTALLSYSTEMET

Gjør om tallet 1101 i totallsystemet til titallsystemet.

……………………………………………………………

# Koder til spørsmål 36

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 13 | 42 | 29,4 | 1,100 | 0,71 |
| | **Ikke riktig svar** | | | | |
| 70 | 11 og 12 | 10 | 7,0 | 0,446 | 0,12 |
| 71 | 14 og 15 | 2 | 1,4 | -0,488 | -0,06 |
| 79 | Andre svar | 15 | 10,5 | -0,579 | -0,20 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 74 | 51,7 | -0534 | -0,58 |

Man kan også lage andre tallsystem enn totallsystemet og titallsystemet. I for eksempel femtallsystemet brukes sifrene 0, 1, 2, 3 og 4.

**Spørsmål 37:** TOTALLSYSTEMET

Hva vil tallet 5 i titallsystemet være i femtallsystemet?

…………………………………………

# Koder til spørsmål 37

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 10 | 11 | 7,7 | 1,485 | 0,43 |
| | **Ikke riktig svar** | | | | |
| 70 | 9 | 4 | 2,8 | 0,428 | 0,07 |
| 79 | Andre svar | 36 | 25,2 | 0,409 | 0,23 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 92 | 64,3 | -0,356 | -0,48 |

---

**Spørsmål 38:** TOTALLSYSTEMET

Gjør om tallet 13 i et femtallsystem til et titallsystem.

…………………………………………………..

---

# Koder til spørsmål 38

| Koder | Svaralternativ | Frekvens | Prosent | Z-skåre | Point-biserial |
|---|---|---|---|---|---|
| | **Riktig svar** | | | | |
| 10 | 8 | 10 | 7 | 1,463 | 0,40 |
| | **Ikke riktig svar** | | | | |
| 70 | 7 og 9 | 5 | 3,5 | 0,943 | 0,18 |
| 79 | Andre svar | 26 | 18,2 | 0,526 | 0,25 |
| | **Ikke svart** | | | | |
| 99 | Blankt | 102 | 71,3 | -0,324 | -0,51 |